

# nature

## SOUND SYSTEM

Making auditory  
connections

## CAUSE FOR OPTIMISM

We are pre-programmed  
to be positive

## BADLY CONTROLLED

Mice were scuttling  
tunnelling technology

# OVER THE BRAINBOW

Multicolour maps of the brain's  
neural connections

NATURE JOBS  
Research & Science

# Who is accountable?

How the responsibilities of co-authors for a scientific paper's integrity could be made more explicit.

**T**he two most notorious frauds of modern science, by the stem-cell biologist Woo Suk Hwang and the physicist Jan Hendrik Schön, both brought into question the responsibilities of co-authors in the oversight of their colleagues' work. But despite the concerns raised after these episodes, there remains a need for a clearer understanding, both within a collaboration and by readers of the eventual papers, of the various contributions made by the authors not only to the research but also to safeguarding its integrity.

One welcome development in transparency was pioneered by the medical journals. Authorship of a paper is justified when a researcher has contributed significantly to the work being described and to the writing or approval of the manuscript. But the traditional publication style is entirely opaque as to which co-author contributed what. Concern about 'honorary authorship' — in which an author is unacceptably included for reasons other than any scientific contribution — and about this lack of transparency has led to the increasing use of statements in papers that specify authors' contributions. Some medical journals require them, and others, including the Nature family, strongly encourage their use and may yet make them compulsory.

Such statements delineate contributions to the work but do not underwrite its integrity. Something more is needed.

It is too glib to state that every co-author of a paper shares full responsibility for its content. A researcher who specializes in the radioactive dating of rock strata cannot necessarily be expected to vouch for a palaeontologist's analysis of fossils within them — especially if the work has been carried out in labs on different continents.

The fact that simple trust may no longer suffice is a sad reflection on recent scientific history, but anything that supports public confidence in research has to be welcomed, provided that its burden is not too great. What follows is a proposal in that direction, on which we invite readers' comments.

We suggest that journals should require that every manuscript has at least one author per collaborating research group who will go on record in a way that collectively vouches for the paper's standards.

Each would sign a statement with reference to *Nature's* publication policies (see [www.nature.com/authors/editorial\\_policies/index.html](http://www.nature.com/authors/editorial_policies/index.html)) as follows:

"I have ensured that every author in my research group has seen and approved this manuscript. The data that are presented in the figures and tables were reviewed in raw form, the analysis and statistics applied are appropriate and the figures are accurate representations of the data. Any manipulations of images conform to *Nature's* guidelines. All journal policies on materials and data sharing, ethical treatment of research subjects, conflicts of interest, biosecurity etc. have been adhered to. I have confidence that all of the conclusions presented are based on accurate extrapolations from the data collected for this study and that my colleagues listed as co-authors have contributed and deserve the designation 'author'."

Principal investigators traditionally bask in the glory of a well-received paper. We are proposing now that they willingly open themselves to sanctions that could be brought to bear should the paper turn out to have major problems.

Misconduct investigators go out of their way to spare anyone apart from the direct perpetrators, but they have indicated concerns over the degree of oversight within collaborations. If the damage to reputations were more widespread in the event of fraud, researchers would be even more fastidious about the data emanating from their labs and the due diligence they would impose. The chances of major frauds, with their disproportionate impact on the reputation of science as a whole, would be diminished.

Readers wishing to offer their views are welcome to contribute to an online discussion at [http://blogs.nature.com/nautilus/2007/10/accountability\\_of\\_co-authors.html](http://blogs.nature.com/nautilus/2007/10/accountability_of_co-authors.html).

**"If the damage to reputations were more widespread in the event of fraud, researchers would be even more fastidious about the data emanating from their labs."**

## In for the cull

A government that asks for independent scientific advice had best be ready to take it.

**T**he question of whether British farmers should be allowed to cull badgers, on the basis that the animals may help spread tuberculosis (TB) among cattle, is perhaps not the most momentous matter on which a government has sought scientific advice. But the mishandling of the issue by David King, the UK government's chief scientific adviser, is an example to governments of how not deal with such advice, once it has been solicited and received.

Back in February 1998, the Independent Scientific Group on Cattle TB (ISG) was set up under the chairmanship of John Bourne, a prominent animal-health specialist, to advise the government department that was responsible for the issue at the time. After much deliberation and the submission of several peer-reviewed papers (such as C. A. Donnelly *et al.* *Nature* **439**, 843–846; 2006), the ISG issued its final report on 18 June this year. Its conclusions were robust: "Badger culling cannot meaningfully contribute to the future control of cattle TB in Britain."

King then proceeded to consider the ISG's report along with, in his words, "other scientific evidence", with the help of five specialists of his choosing. On 30 July he gave his report to the secretary of state with a startlingly different conclusion. "Removal of badgers," it states,

“should take place alongside the continued application of controls on cattle.” This report was made public on 22 October.

Last week, King was rightly criticized by scientists and members of parliament for seeming to go back on the ISG’s advice, which the government had itself sought. Badger culling is a politically fraught issue in Britain, pitching farmers against the equally passionate and vocal animal-protection lobby. King’s motives remain unknown but his actions are likely to encourage speculation that his report was written to please the farmers.

In many instances, it is likely that political factors will ultimately overrule scientific ones when a government takes a decision in a contentious field. If this is the case, then surely it would be better not to seek independent scientific advice that will inevitably be ignored. There are countless examples — the planned replacement of the Trident nuclear submarine arsenal, for instance — in which the UK government had no intention of taking independent advice, and so had the good manners not to ask for it.

In the United States, researchers are accustomed to treating the process that feeds scientific advice into the government with some suspicion. The latest incident, in which presidential science adviser John Marburger stands accused of interfering with testimony on climate change and public health first submitted by the Centers for Disease Control and Prevention, merely reinforces this atmosphere (see page 8).

But in Britain, scientists have enjoyed a better relationship with their government and — prior to the badgers episode — little evidence has come to light of advisory recommendations from scientists being cooked or spun to match the government’s intentions.

On 24 October, Bourne and King were called to account for what had happened at a meeting of the House of Commons select committee on Environment, Food and Rural Affairs. Bourne was visibly annoyed, and described King’s report as “hastily written” and “superficial”. Rosie Woodroffe, an expert on conservation biology at the University of California, Davis, and an ISG member, said that the King report was riddled with “small mistakes”. In those circumstances, King’s insistence that “the conclusions in my report are not very different from those that the ISG reached” ring hollow.

It would be a good idea if the Department for Environment, Food and Rural Affairs, which is now responsible for the matter, based its policy on the unfettered advice offered by Bourne’s committee. This would be deeply appreciated not just by the badgers, but by scientists in all spheres who choose to participate in painstaking advisory processes in the earnest belief that their advice will actually make a difference to government policy. ■

**“Political factors will ultimately overrule scientific ones when a government takes a decision in a contentious field.”**

## Because it's there

An Asian Moon race is neither particularly worrying nor especially inspiring.

In the 1960s and 1970s, the United States and Soviet Union mounted dozens of missions to the Moon, orbiting it, crashing into it and landing softly on it. They even went so far as to return samples from it, either with a little help from some humans on hand or, in the Soviet case, without. Subsequently, neither spacefaring power touched the place for almost 20 years. In part this is because their race, such as it was, had ended. It was also because planetary scientists were far more interested in exploring other places. The Moon had a distinct been-there-done-that aura.

But for aspiring nations that have neither been there nor done that, the Moon has a great advantage over other objects of celestial study. Although only moderately interesting, it is very close and relatively easy to reach. So in 2003 it was the obvious target for Europe’s SMART-1 mission, which tested a new sort of rocket propulsion. And it is currently the destination of choice for others seeking to develop their spacefaring prowess.

In September, Japan finally followed a very small lunar mission launched in the 1990s, Hiten, with a much larger and more ambitious one, SELENE. October saw the launch of Chang’e-1 from China (see page 12), timed to coincide with the ‘re-election’ of Hu Jintao as leader of the Communist Party — a piece of celestial theatrics well in tune with the spirit of technocratic command and control that characterized the original Moon race. Next year will see the launch of India’s

Chandrayaan-1 and America’s Lunar Reconnaissance Orbiter, the heavyweight of the current crop.

It is easy to exaggerate the extent to which this constitutes a new Moon race. National rivalries and prestige definitely play a part in some of these programmes: China’s, in particular, is both touted by the government and appreciated by the population as evidence of national accomplishment and ambition. But the idea sometimes floated that this activity reflects a new perception of some sort of value in the Moon itself is wide of the mark.

Although there are interesting scientific questions about the Moon, few, if any, are of the first order. And despite some hype to the contrary, the Moon’s potential as a source of raw materials for Earth’s consumers is ludicrously constrained. There is nothing there worth the cost of bringing back to Earth (not even helium-3, a fuel of unknown utility to a second or third generation of fusion reactors of unknown feasibility). The Moon’s potential as a resupply station for spacefarers visiting other places is also poor, although perhaps not entirely negligible.

That doesn’t mean that the current spate of missions to the Moon is worthless. One learns by trying, and the Moon is a good test bed for mastering the arts of planetary exploration. The same applies, further down the line, to the far more resource-intensive business of sending humans; if you feel you must send humans elsewhere, it is a conveniently near at hand and well-characterized destination.

But the only prize to be won in any race that ends with humans yet again walking on the Moon is global recognition that you have managed to do what was accomplished to little lasting effect back in the days of flower power. It remains unclear that such recognition is worth the already stretched resources of India or China — or of any other nation. ■

# RESEARCH HIGHLIGHTS

## MOLECULAR NEUROBIOLOGY

### Active resilience

*Cell* **131**, 391–404 (2007)

A molecule known as BDNF may provide a pivotal distinction between people who succumb to conditions such as depression or post-traumatic stress disorder and the majority who do not.

Eric Nestler of the University of Texas Southwestern Medical Center in Dallas and his colleagues had previously identified mice that avoided social contact with cage-mates after repeated 'social defeat' — brought about by forced encounters with more aggressive mice.

The researchers found that resilience to stress in this social-defeat model is an active process in which stress-induced firing in part of the brain's neuronal circuitry associated with reward and drug addiction is suppressed. This lowers levels of brain-derived neurotrophic factor (BDNF), and mice with a natural variation in their BDNF gene were resilient to stress. The researchers also found high levels of BDNF in a region associated with reward in post-mortem brains of people with a history of depression.

## ECOLOGY

### Smelly fish

*Proc. R. Soc. B* doi:10.1098/rspb.2007.1283 (2007)

Fish that have been exposed to even very low levels of a common pollutant are shunned by their conspecifics, according to research by Ashley Ward at the University of Sydney and his co-workers. The effect may put the tainted fish at increased risk of predation.

The research looks at the effect of low levels of 4-nonylphenol, a surfactant sometimes found in shampoos and soaps, on schooling in juvenile banded killifish (*Fundulus diaphanous*, pictured below). Unexposed fish were unwilling to school with fish that had been exposed for as little as an hour to levels commonly found in sewage outflows.

### Extraordinary sound

*Nature Phys.* doi:10.1038/nphys774 (2007)

Light squeezed through a hole of smaller diameter than its wavelength can emerge in a collimated beam when the hole is set in a periodic array of indentations. This counter-intuitive phenomenon is known as "extraordinary optical transmission". Now researchers in Spain have shown that an analogous effect can be achieved with sound waves.

Using their understanding of the role of surface plasmons in the optical case, Francisco Garcia-Vidal at the Autonomous University of Madrid and his colleagues designed systems of holes and grooves in which acoustic surface waves have a similar effect, producing a beam of sound through cunningly contrived interference and reinforcement. The authors think this effect might be applied in non-destructive materials testing and medical ultrasound.



## MICROBIOLOGY

### A time to die

*Science* **318**, 652–655 (2007); *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0704256104 (2007)

Bacteria communicate with each other to a greater extent, and in more ways, than was once thought. Hanna Engelberg-Kulka of the Hebrew University in Jerusalem and her colleagues report that *Escherichia coli* can send out messages that encourage their neighbours to kill themselves. The 'extracellular death factor' involved turns out to be a chain of five amino acids that seems to be produced by the degradation of a metabolic enzyme, glucose-6-phosphate dehydrogenase.

The authors suggest that the ability to coordinate cell death may allow bacteria to release nutrients from a proportion of individuals when the population is under stress, or to defeat phage attacks by means similar to a 'scorched-earth' policy.

Meanwhile, Frances Arnold at the California

Institute of Technology in Pasadena, Ron Weiss at Princeton University, New Jersey, and their colleagues report engineering a system into populations of *E. coli* that allows a specific pattern of gene expression only when both populations are present in sufficient numbers. They suggest that such systems could be used to engineer specific responses in artificial biofilms.

Intriguingly, three years ago these labs engineered a 'neighbour killing' system into bacteria quite similar to that now being reported (L. You *et al.* *Nature* **428**, 868–871; 2004). This may be the first example of synthetic biology anticipating a function only later revealed in nature.

## RELATIVISTIC PHYSICS

### Hot at any speed

*Phys. Rev. Lett.* **99**, 170601 (2007)

According to the theory of special relativity, measurements of length and time differ depending on the relative velocity of the observer and the thing being measured. Einstein and other prominent physicists believed that the measured temperature of a gas would also depend on the speed of the observer.





Work by Peter Hänggi of the University of Augsburg in Germany and his collaborators contradicts those early calculations. The group's one-dimensional models of particles in a gas show that the same temperature will be observed regardless of the observer's speed. The team admits, however, that this may not be true of two- or three-dimensional gases, and believes that further study is needed.

## CLIMATE CHANGE

### Irreducible sensitivity

*Science* **318**, 629–632; 582–583 (2007)

Decades of intensive work on climate change have done little to narrow the range of uncertainty over Earth's climate sensitivity — the amount of global warming to be expected under carbon dioxide levels twice those of the pre-industrial world, which the Intergovernmental Panel on Climate Change puts at between 2 °C and 4.5 °C. Nor can a substantially better estimate be expected in the future, according to Gerard Roe and Marcia Baker of the University of Washington in Seattle.

Their mathematical analysis stresses that normally distributed uncertainties in the feedbacks associated with climate processes make ruling out a long tail of high sensitivities unrealistic. In an accompanying Perspective, Myles Allen and David Frame at the University of Oxford, UK, argue that this is not a problem for setting policy, because the concentrations of carbon dioxide associated with specific warmings can be much better constrained.

## CHEMISTRY

### Clicking without copper

*Proc. Natl Acad. Sci. USA* **104**, 16793–16797 (2007)

Proteins can be quite easily labelled for studies in living cells, but other types of molecule are trickier to track. One possibility is to use a two-step technique: first get the cell's own machinery to add a chemical 'socket' to the molecule of interest, then add a trackable reagent that fits easily into that socket. 'Click' chemistry — so named for the ease of assembly — might be helpful for this second step, but normally depends on toxic copper compounds.

Now Carolyn Bertozzi and her colleagues at the University of California, Berkeley, report using a cyclooctyne onto which fluorine atoms have been added to carry out the same sort of click reaction without the need of copper. Their system can be used to track sugar molecules in and on living cells for minutes or for days.

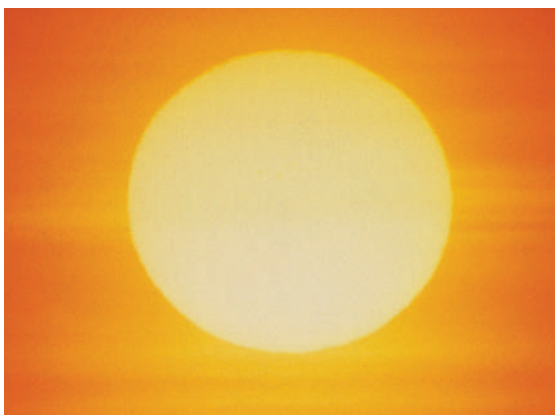
## PLANETARY SCIENCE

### Identical twins

*Astrophys J.* **669**, L89–L92 (2007)

Astronomers have identified a star that is in many ways indistinguishable from the Sun.

Peruvian astronomers Jorge Meléndez of the Australian National University, and Iván Ramírez at the McDonald Observatory of the University of Texas in Austin report that the parameters of HIP 56948, one of four 'solar twins' they have been studying, are exactly the same as the Sun's, within the constraints of observational accuracy. Unlike previous solar twins, this star — which resides 200



S. MCCARTNEY/SPL

light years away in the constellation of Draco — shares the Sun's low lithium content. In addition, HIP 56948, like the Sun, has no accompanying 'hot Jupiter' in close orbit.

Such solar twins are useful for various calibrations, the authors say; they are also of potential interest to researchers involved in the search for extraterrestrial intelligence.

## CHEMICAL PHYSICS

### Nano steams ahead

*Phys. Rev. Lett.* doi:10.1103/PhysRevLett.99.178103 (2007)

Nanoparticles may exhibit chemotaxis — directed motion along a gradient of chemical concentration previously seen at the microscale only for living cells.

Ayusman Sen, Darrell Velegol and their colleagues at Pennsylvania State University have shown previously that nanorods made with platinum at one end and gold at the other move through a solution of hydrogen peroxide, owing to the catalytic production of oxygen gas. Now they report that in a peroxide gradient the rods — 370 nm wide and 2 µm long — move 'uphill' towards the fuel source, because the increase in speed at higher peroxide concentrations biases their movement. This offers a simple way to power and guide nanoscale objects.

## JOURNAL CLUB

James E. Ferrell

Stanford University School of Medicine, California, USA

### A systems biologist encourages modelling by the millions.

In a typical modelling study, we write down equations, solve them, and see whether they account for known data. If they do, we claim to understand some bit of biology. One huge caveat is that many other models might have matched the data just as well.

Researchers from Peking University in Beijing and the University of California, San Francisco, have devised a satisfying way of dealing with this problem (W. Ma *et al.* *Mol. Syst. Biol.* **2**, 70; 2006).

Their starting point was epithelial patterning in the fruitfly *Drosophila*. During embryogenesis, a system known as the 'segment polarity network' generates repeating stripes of gene expression. The stripes are initially fuzzy and later become sharp. Ma *et al.* set out to see what simple gene circuits were best suited to this sharpening process.

They formulated differential-equation models for about 14 million ways of connecting two or three segmentation genes, then randomly chose 100 sets of parameters that defined the strength of the interactions for each gene. They then carried out computations for each combination to determine which of them converted fuzzy stripes into sharp ones.

Many topologies worked for at least one parameter set. But only a fraction worked for more than one or two. Interestingly, the most robust topologies were all variations on the same design — each had three sub-circuits, one 'stripe generator' motif and two bistable 'response sharpeners'. These findings give hope that complex networks may be decomposed into modular sub-circuits with understandable functions.

Comprehensively examining millions of models is a lot of work, but is not impossible. And, as Ma *et al.* show, it can yield important insight that could not have been derived from studies of one or two.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

## SPECIAL REPORT

# Model behaviour

The brain is no longer the black box it used to be, and neuroscientists are starting to put new knowledge to good use, developing better animal models for psychiatric disorders. **Alison Abbott** reports.

Dangle a mouse by its tail, and it will wriggle and strain to escape before eventually recognizing the hopelessness of its situation. Measure the time it takes to abandon thoughts of helping itself, and you have one of the classic animal tests for depression.

Except it's not, says Laurence Tecott, a research psychiatrist at the University of California, San Francisco. "We can't say that that mouse is depressed, and we can't say you would be if you were strung up by your tail," he says. The reason we have not seen a genuinely new class of drug in psychiatry for 50 years, he asserts, is largely because animal models are woefully inadequate representations of human-specific disorders.

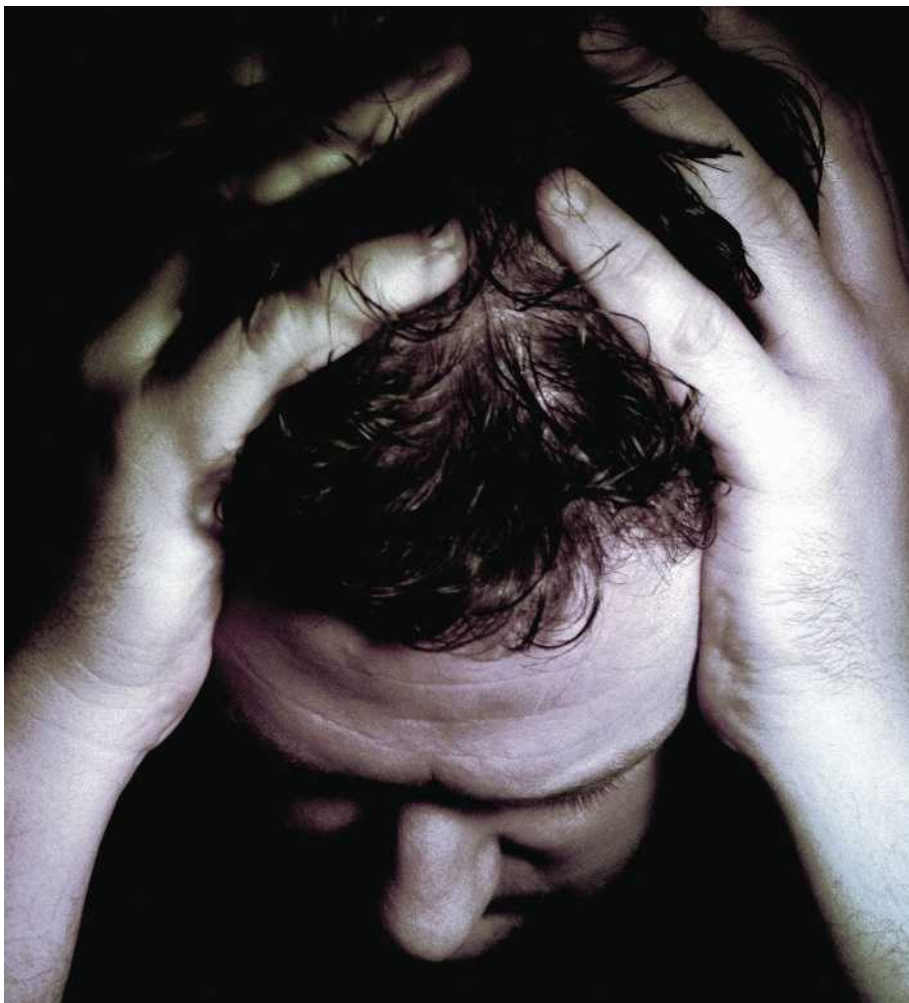
You'll hear the same story from many others. But things are not as hopeless for scientists as they may seem for the dangling mouse; some recent papers offer tantalizing hints of a way forward. "No one is going to create a mouse model of suicidality," says Eric Nestler, a neuroscientist at the University of Texas Southwestern Medical Center at Dallas. "But sensible models of important aspects of the neurobiology underpinning psychiatric disorders are just around the corner."

Classical animal tests for psychiatric disorders are based on responses to clinically proven drugs. What the tests don't necessarily do, however, is reflect the cause or the biological basis of the disorder they are supposed to mimic. Most researchers agree that it's time to apply recent findings about the human brain to creating more useful mouse models — often by deleting, adding or mutating candidate susceptibility genes.

For instance, scientists have already modelled a core symptom — abnormal social relations — of several autism-related disorders<sup>1,2</sup>. And schizophrenia, a disorder believed to result in part from faulty neurodevelopment, has also been a focus of recent research, because a particularly strong candidate gene has been identified — that encoding the protein DISC-1.

## Head start

Psychiatrist Akira Sawa, from the Johns Hopkins University in Baltimore, Maryland, for example, has created transgenic mice with a disrupted *Disc-1* gene and used brain scanning to show that such mice have enlarged brain ventricles — especially on the left side — as is usually seen in humans with schizophrenia<sup>3</sup>.



VICTOR HABBICK VISIONS/SPL

**Better animal models of human psychiatric disorders will be vital to developing new types of treatment.**

"This brain anatomy is a very objective characteristic of schizophrenia," says Sawa. "And being able to reproduce it in mice with the candidate association gene further supports the hypothesis that DISC-1 is involved in schizophrenia." His lab is now looking at proteins that interact with DISC-1 in key biochemical signalling pathways, and how environmental factors such as stress may affect the system.

Alcino Silva, a research psychiatrist at the University of California, Los Angeles, has gone a step further. His team has created a 'conditional mutant' mouse, in which the *Disc-1* gene can be switched on or off at will. He found that activation of the gene on just one specific day

— the seventh after birth — was sufficient to cause a range of symptoms analogous to those seen in schizophrenia later in life<sup>4</sup>. The mice developed specific types of memory deficit and depressive-like behaviours, and were less sociable with other mice. In addition, the hippocampus, a part of the brain that is important in memory, had notably different anatomy at the cellular level. "When we went back to look at natural mutations in the gene in humans, we found that the mutation was correlated with reduced sociability," says Silva. "This may not be an accurate model of schizophrenia, but it is telling us a lot about the disease."

Few, if any, scientists want or expect to recreate



**SOCIETY FOR  
NEUROSCIENCE BLOG**  
Read our diary reports from  
this year's meeting  
[http://tinyurl.com/  
2cbq9d](http://tinyurl.com/2cbq9d)



A mouse mutant (left) for *Disc-1*, a gene implicated in schizophrenia, quickly gives up when forced to swim.

an entire psychiatric disorder in a mouse. This would be beside the point, not least because current clinical diagnostic criteria, such as those enshrined in the *Diagnostic and Statistical Manual of Mental Disease*<sup>5</sup> (DSM-IV for the fourth, and current, edition) are also considered inadequate. Most neuroscientists and psychiatrists agree that they add to confusion by giving names to clusters of symptoms that overlap with each other. "Schizophrenia and bipolar disorder, as defined by DSM-IV, share many symptoms, as do anxiety and depression," says neuroscientist Steve Hyman, provost of Harvard University in Cambridge, Massachusetts. "The emerging models focus on biomarkers of specific symptoms, rather than a DSM label."

### Circuit breaker

Particularly valuable biomarkers, says Hyman, are the neural circuits relevant to disease that have been identified in people using techniques such as functional magnetic resonance imaging (fMRI), which measures changes in oxygenation of blood flowing through the brain. For example, a circuit that runs between the portion of the cortex at the front of the brain, and the striatum and the thalamus deeper in the brain, has been associated with obsessive-compulsive disorder. The symptoms of this disorder can sometimes be alleviated by drugs such as selective serotonin reuptake inhibitors (SSRIs), which enhance the effects of the neurotransmitter serotonin.

Guoping Feng, a neurobiologist at Duke University Medical Center in Durham, North Carolina, hit this target in mice, although not by design. Six years ago, he knocked out a gene in mice called *Sapap3*, which encodes a key protein involved in regulating receptors for the

neurotransmitter glutamate. He noticed that the genetically altered mice groomed themselves until they bled.

At first, he had no idea why. He put things together when he remembered that SAPAP3 is the only member of the SAPAP protein family that is highly expressed in the striatum. And he thus ended up with the first model in which specific disturbances in this neural circuit are associated with a compulsive-like behaviour<sup>6</sup>. Selectively expressing the missing gene in the striatum alleviated the abnormal behaviour, as did giving the mice SSRIs. "Drug companies have contacted us about testing small molecules they have in their cupboards which regulate glutamate receptors," says Feng.

Such research is also stimulating companies to rethink their approach to testing drugs.

During recent decades, the available arsenal of psychiatric drugs has been fine-tuned so that fewer patients have severe side effects. But there has been little improvement in efficacy. SSRIs don't work in one-third of patients with depression, for example, and antipsychotic drugs have little impact on the debilitating cognitive deficits of schizophrenia, even though they control disruptive symptoms such as hallucinations. "Current tests continue to pick up drugs that work through the same old mechanisms," says Michael Spedding, director of research at the pharmaceutical company Servier in Paris, "which is why most schizophrenia drugs target the dopamine system, even though that's not good enough therapeutically." Spedding is now trying to develop surrogates in animals

that involve the relevant neural circuits.

Mark Tricklebank, head of science at Eli Lilly's Centre for Cognitive Neuroscience in Windlesham, Surrey, UK, says that many industrial pharmacologists are starting to change their approach in this way. His group, for instance, is developing an oxygen electrode to measure blood-oxygenation levels directly in particular areas of the brain in rodents, as a surrogate of fMRI signals seen in humans.

But it will be some years before the new approach becomes mainstream. "Trying to make drugs is a ferociously complex process," says Paul Chapman, head of GlaxoSmith-Kline's Centre for Research in Cognitive and Neurodegenerative Disorders in Singapore, "and tolerance to risk varies from company to company". The stakes are huge. A sixth-generation SSRI, Lexapro (escitalopram), has so far made almost US\$2 billion for New-York-based Forest Laboratories; by contrast, the last major attempt to launch a genuinely new psychiatric drug, based on the neurotransmitter substance P, failed to show efficacy in several clinical trials for depression.

But it is also early days scientifically. Neuroscientists are quick to point out that their new knowledge about the brain is still preliminary and sketchy. Current candidate association genes are just that — candidates. And the interesting neural circuits are not proven disease mechanisms, just promising correlates.

In the meantime, Tecott is preparing to unveil a new approach to assessing mouse behaviours. It's a holistic approach, whereby the 'lifestyle' of a mouse is assessed in its home cage — a very particular home, where every

move is monitored in space and time, along with every lick, every milligram of food eaten, and every snooze. The data can then be mined for information about patterns of behavioural change in the short and long term, for example during the female reproductive cycle, or as a

**"Current tests  
continue to pick  
up drugs that work  
through the same  
old mechanisms."  
— Michael Spedding**

consequence of drug administration.

"Looking at spontaneous behaviour is like looking at brain function on display," says Tecott. "This will help open up ways of finding new psychiatric drugs at last."

**Alison Abbott**

1. McGill, B. E. et al. *Proc. Natl. Acad. Sci. USA* **103**, 18267–18272 (2006).
2. Tabuchi, K. et al. *Science* **318**, 71–76 (2007).
3. Hikida, T. et al. *Proc. Natl. Acad. Sci. USA* **104**, 14501–14506, (2007).
4. Li, W. et al. *Proc. Natl. Acad. Sci. USA* (in the press doi:10.1073/pnas.0706900104).
5. *Diagnostic and Statistical Manual of Mental Disorders* 4th Edn (Am. Psychiatric Press, Virginia, 2002).
6. Welch, J.M. et al. *Nature* **448**, 894–901 (2007).



## SCORECARD

**Biodiesel buses**

Passengers in the UK town of Kilmarnock are getting a 20p (41¢) discount on their bus fare if they bring their used cooking oil, which is converted into biodiesel for the fleet.

**Solar-powered cars**

Competitors from Stanford (below) and Michigan universities were left red-faced after both teams' cars suffered accidents in the World Solar Challenge race in Australia.



PANASONIC WORLD SOLAR CHALLENGE

## ON THE RECORD

**“Science changes with every generation and with new discoveries and God doesn’t. So I’ll stick with God if the two are in conflict.”**

Republican presidential candidate Mike Huckabee fails to cover all the bases in the science-religion debate.

## ZOO NEWS

**Deep time**

A clam dredged up off the Icelandic coast that seems to have been there for as much as 410 years has been hailed as the longest-lived animal ever discovered. Of course, it might have lived even longer if it hadn’t been dredged up and given to age researchers...

## NUMBER CRUNCH

**33%** of Swedes have visited one of the country’s three nuclear power plants since they were opened to tourists 35 years ago.

**80%** of Swedes are “comfortable” with the country’s continued use of nuclear energy.

**275** safety actions are under way at the Forsmark nuclear plant, after an emergency shutdown in July 2006.

Sources: BBC, The Stanford Daily, Univ. Michigan Solar Car Blog, Salon.com, Reuters

# White House adviser linked to cuts in climate-change report

The chief science adviser to President George W. Bush came under fire last week for his role in watering down congressional testimony on the health effects of global warming.

Julie Gerberding, director of the federal Centers for Disease Control and Prevention in Atlanta, Georgia, had briefed senators on the subject at a hearing on 23 October. As required, she submitted her written testimony 24 hours in advance to the White House — which then proceeded to chop it in half. Detailed comments from the office of John Marburger, the president’s science adviser and director of the Office of Science and Technology Policy (OSTP), may have played a part.

Six pages of missing material focused on a range of potential public-health impacts related to climate change. These included the effects of heat waves, air pollution, extreme weather and infectious diseases. The original testimony also indicated that children, the elderly and the poor would bear the brunt of the impact, and suggested that these problems “remain largely unaddressed”.

Officials at the Centers for Disease Control maintain that Gerberding’s testimony was accurate and complete, especially when combined with what she told the committee verbally. But congressional Democrats and advocacy groups immediately denounced the White House, citing a history of allegations that the administration has misrepresented science in order to promote its own policies.

“Considering the track record, this administration should be bending over backwards to allow critical scientific evidence to reach the public,” says Tim Donaghy of the Union of Concerned Scientists, an advocacy group based in Cambridge, Massachusetts.

Marburger insists that his office did not seek to cut the report, but instead made “substantive and constructive comments and suggestions”. Those comments were passed to the White House Office of Management and Budget, which coordinates such agency reviews. Owing to time constraints, the budget office elected to strike the sections rather than revise the

testimony, according to OSTP officials.

Donaghy acknowledges that Marburger’s comments are technically accurate, but he says the administration had a responsibility to correct the errors rather than cutting the meat out of the testimony. “The topic of the hearing was climate change and public health,” he says. “It’s not too much to ask to have them submit testimony on that particular topic.”

The budget office referred all questions on the matter to the White House press office, which declined to discuss the issue.

At the time *Nature* went to press, the OSTP had said that it would not release the full version of its suggested edits, as such reviews were an “internal process”. But in a written statement, Marburger offered several examples of what he called an “overall lack of precision” in the proposed testimony, and cited what he claimed were technical conflicts with

the latest assessment by the Intergovernmental Panel on Climate Change (IPCC).

For instance, Marburger flagged a statement indicating that global warming is expected to alter the “frequency, timing, intensity and duration of extreme weather events such as hurricanes and floods”. But researchers have made links between global warming and hurricanes primarily for storm intensity, and even on that point there remains considerable debate.

Marburger also said that the report improperly characterized certain global impacts as domestic impacts, and failed to acknowledge that, in some cases such as agriculture, the impact in the United States might be beneficial. He also noted that the effects on mental health anticipated by the IPCC relate to post-disaster trauma, not the anticipation of global warming as stated in the testimony.

Senator Barbara Boxer (Democrat, California), who chairs the Senate Committee on Environment and Public Works that called Gerberding to testify, has asked the White House for copies of all records related to the testimony. She calls Marburger’s explanation a “lame defence”.

**Jeff Tollefson**

**See Editorial, page 1.**



John Marburger says global-warming testimony ‘lacked precision’.

M. EULER/AP



# Sponsor hunt divides geosciences union

Talks about approaching the oil company ExxonMobil as a possible meeting sponsor are raising tempers at the European Geosciences Union (EGU) in Strasbourg, France. Some scientists fear that sponsorship from the company would damage the integrity of the organization.

The EGU council argued over its options at a meeting on 12 October, but made no decision as to which companies should be approached. But *Nature* has learned that ExxonMobil is likely to be on a list of possible sponsors that the council has asked John Ludden, the EGU's vice-president, to compile in the upcoming weeks.

The 25-member council did agree that it would need to discuss and agree on potential sponsors before they were approached. If it could not reach a consensus, an ethics committee — the members of which will be appointed shortly — would weigh in.

ExxonMobil, which last year reported record earnings of US\$39.5 billion, donates tens of millions of dollars every year to universities and scientific organizations, including the American Geophysical Union (AGU) in Washington DC. However, the company has been criticized in the past for also funding climate sceptic groups that critics say attempt to cloud the scientific understanding of climate change and delay action against global warming.

The company has now stopped funding some of these groups, such as the Competitive Enterprise Institute in Washington DC. And its 2006 corporate citizen report states that “it is prudent to develop and implement strategies” to deal with rising greenhouse-gas emissions.

Unlike its American sister organization, the EGU — which launched in 2002 as a merger of two other groups — has not accepted funding from private sponsors. ExxonMobil did fund a student lunch during the EGU's 2006 meeting in Vienna, Austria, but the event, which was to help students make contact with potential employers, was not an official part of the conference.

The EGU's annual meetings, which attract

some 8,000 scientists, generate most of the group's income. But the high registration fee — up to €400 (US\$577) next year — is not affordable for many scientists and students from poorer countries. “We're not going to go bankrupt,” says Ludden. “But in order to attract more scientists from Africa and eastern Europe we really should start looking for suitable sponsors.”

Like many geologists, for whom oil companies are the main non-academic employers, Ludden says that he has no problems with accepting funds from the oil industry.



ExxonMobil is one of several companies being considered as a sponsor for scientific meetings.

He hopes that introducing sponsors will raise several hundred thousand euros per year for student sponsorship, teachers' education and for underwriting the costs of new journals.

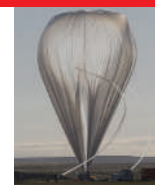
Tim Killeen, president of the AGU, says that issues of sponsorship there “have been a topic of lively discussion for many years”. In the European group, dissent is led mainly by climate scientists and those outside the solid-earth community. “Although we should be open to various types of interaction with industry, we shouldn't approach certain companies just because they're free with money,” says Hermann Held, chair of the EGU's division for energy, resources and the environment.

If sponsorship were to move forward, it would be unlikely to be in place for the 2008 meeting in Vienna, says Ludden.

“It's a delicate issue,” says Gerald Ganssen, the EGU's president. “We'll take all necessary means to solve this internally to avoid the risk of splitting the EGU.”

Quirin Schiermeier

**“We should be open to various types of interaction with industry.”**



### BALLOON TAKES TELESCOPE TO GREAT HEIGHTS

Test flight uses balloons for cheap astronomy.

[www.nature.com/news](http://www.nature.com/news)

C. CALVIN, UCAR

# Personalized genomes go mainstream

When scientists released a draft of the human genome sequence six years ago, they said the data belonged to all of us — but until now, they have been the only ones able to play with the data therein.

That's now starting to change. In the next few months, two Silicon Valley start-ups will start giving customers a peek at their genomes for a few thousand dollars a pop. A firm in Massachusetts is also offering people the opportunity to buy the whole sequence of their genome for an unspecified amount. And close behind these firms are corporations, such as Google, that are developing ways to store, analyse and profit from health information — including genetic data.

The first start-up, 23andMe of Mountain View, California, plans to launch before the end of this year, and the second, Navigenics of Redwood Shores, hopes to do so by spring 2008. Both companies say that they will genotype millions of regions in customers' genomes, called single nucleotide polymorphisms or SNPs, which have been linked to a handful of diseases and non-medical traits, such as earwax consistency. They then sell that information back to the customer.

Navigenics will focus on medical conditions, says its chief scientific officer, Dietrich Stephan, a human geneticist at the Translational Genomics Research Institute in Phoenix, Arizona. It will use information from scientific studies to estimate composite risk factors for diseases based on each customer's SNPs, he says. It also plans to provide genetic counselling to help customers interpret these risk factors. The firm says that it will conduct long-term studies on how well those predictors work.

By contrast, 23andMe will analyse not just



Commercial personal genome services could help people to understand more about their family medical history.

medical information, but also traits not necessarily linked to disease. The firm will introduce a social networking component to genomics by allowing customers to link their data with others, such as family members, say advisers. It is also considering providing researchers with access to the data, they say.

Advisers familiar with the firms say that there is a distinct difference between them. "Navigenics might be the white coat and stethoscopes approach, and 23andMe might be about family and ancestry and who you inherited something from," says genome scientist George Church from the Massachusetts Institute of Technology in Cambridge. Church is an adviser for 23andMe, has had discussions with Navigenics and is also advising Knome, the Cambridge-based personal genome company that offers to sequence the whole genome. "It's like the difference between IBM in the 1960s, which was about black ties and white shirts and shiny black shoes, and Google in the present time, which is

a little more playful," he says.

Indeed, one of 23andMe's founders, Anne Wojcicki, is married to Google founder Sergey Brin, and Google has invested millions in her company. Google is developing its own suite of health tools to allow users to personalize and share health information, and many speculate that 23andMe will feed its data to Google.

The prospect of Google working with health data has raised concerns about privacy. Google typically targets advertisements to customers from information they enter into search engines and from the content of their e-mail messages, raising the issue of how it would handle personalized health records and genetic data. Few laws govern such use of health or genetic information — a sobering fact, given that any individual can be identified by 100 SNPs or fewer (Z. Lin, A.B. Owen and R. B. Altman *Science* **305**, 183; 2004).

But many customers might not be bothered by the privacy issue, says John Halamka, chief information officer for Beth Israel Deaconess Medical Center in Boston, Massachusetts, and a Google adviser. "Every one of us has a different threshold for privacy," he says.

## Share and share alike

In fact, Halamka will be sharing some of his genetic data with the world next month as part of the Personal Genome Project led by Church. The project is releasing genetic and other types of data on ten volunteers as an experiment to test how such data are used, or misused, by participants and the public (see 'Genetic test claims AIDS insight').

Church points out that hundreds of thousands of people have already paid for services that analyse the ancestral origins of their DNA. So companies such as 23andMe and Navigenics are likely to make money in the short term.

But whether the firms stick around will depend on whether they prove their worth, says Kathy Hudson, director of the Johns Hopkins Genetics and Public Policy Center in Washington DC. Hudson points out that the power of SNP-based tests is still very limited, so a whole-genome scan can't tell you much that you can't find out another, cheaper way.

"My real question is: why are you getting a genetic test to find out if you have sticky earwax?" Hudson asks. "I don't need a genetic test; I've got my Q-tips."

Erika Check Hayden

## Genetic test claims AIDS insight

"Excited and relieved are the first words that come to mind." So says a "satisfied customer" on the website of HIVmirror, a company based in Philadelphia, Pennsylvania. The firm says that it "tests people for a gene that can slow HIV down from becoming AIDS".

The test looks for particular variants of two genes, *CCR5* and *CCR2*, used by HIV to

gain entry into human cells. Studies have found that on average, people with specific variants in these genes are less likely to progress quickly to AIDS than are those with other variants. But as the studies have looked at average data over large groups of patients, they cannot make definitive predictions about any individual, raising questions

about how HIVmirror's test results will be used — or misused — by customers.

HIVmirror declined to comment to *Nature*. But its test does not require approval from the US Food and Drug Administration because regulation of genetic tests lags far behind the science — a problem that is likely to get worse as such tests become more available. **E.C.H.**



# China blasts off on lunar exploration

China's Moon exploration programme has kicked off with the launch of its first lunar orbiter. Launched on 24 October from Sichuan province in southwest China, the orbiter is a small step in an ambitious, and not yet official, plan to send astronauts to the Moon sometime after 2020. But its significance lies in China's hopes of becoming the fifth power to put a probe into lunar orbit.

"It's a technological confidence-builder," says Joan Johnson-Freese, a China space analyst at the Naval War College in Newport, Rhode Island.

As is often the case when China flexes its muscles, the world has taken notice. Analysts say China's space programme is spurring competition among its Asian neighbours, even as it seeks assistance and collaboration from Europe. Meanwhile, the United States, wary of military competition in space following China's anti-satellite weapons test in January, is monitoring the programme carefully.

The 2.3-tonne, US\$190-million Chang'e-1 spacecraft — named after a mythical Chinese goddess whom legend says flew to the Moon — rode on a Long March 3A rocket that blasted off from the Xichang launch centre. If all goes to plan, by 5 November the probe will have parked itself in an orbit around the Moon. It will then begin a year-long geological and mineralogical analysis as it inspects the surface for future landing sites.

Chang'e is carrying a camera and an altimeter to build a three-dimensional map of the lunar surface, spectrometers to determine the Moon's elemental composition, a radiometer to map soil depths by measuring reflected microwaves, and detectors to monitor the solar wind.

That will set the stage for Chang'e-2, which aims to land a rover on the Moon by 2012. The rover would transmit video footage and analyse soil samples. Scientists in Shanghai are developing a nuclear-powered prototype that resembles NASA's current Mars rovers Spirit and Opportunity, although it is not clear if China's space agency will select this design. A possible subsequent Chang'e-3 mission would gather lunar rock samples and return them to Earth.

China's space missions are organized as five-year plans, each designed to make significant technological advances over the previous one. "By Western standards,



China's lunar probe lifts off from the Xichang launch centre, aiming to reach the Moon by 5 November.

it's painfully slow," says Phillip Clark, head of the Molniya Space Consultancy in Hastings, UK, who has studied the history of China's space programme.

Johnson-Freese notes that the Chang'e missions are exploratory science missions managed by the nominally civilian China National Space Administration, as opposed to China's manned missions, which are under the control of the People's Liberation Army. In 2003, the Shenzhou 5 mission carried the nation's first

astronaut, Yang Liwei, into orbit, making China the third country to do this after the United States and the Soviet Union. Two astronauts flew aboard Shenzhou 6 in 2005. A spacewalk is the next milestone, planned for next year.

Eventually, the capabilities of the two programmes are expected to merge. But there have been few official pronouncements from the top levels of government. A deputy commander for the Shenzhou programme has said China could have people on the Moon by 2020; Clark thinks such a date is optimistic, but possible. Top Communist Party officials have avoided committing to a specific date, probably for fear of making it too much of a competition with the United States, says Vincent Sabathier, director for space initiatives at the Center for Strategic and International Studies in Washington DC. "They're trying to avoid this race," he says.

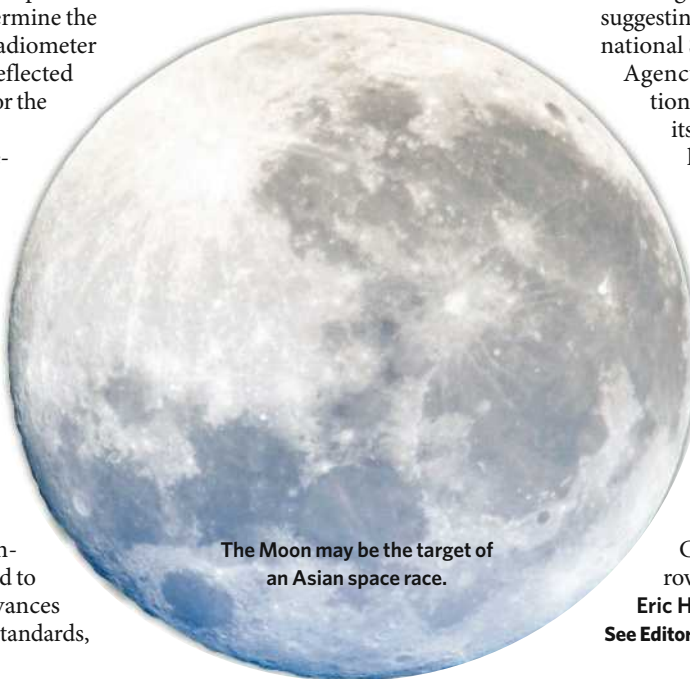
Yet in September, NASA administrator Michael Griffin evoked the language of the old US-Soviet space race when he said China might very well beat America in sending astronauts back to the Moon. Some observers, including Johnson-Freese, say this was perhaps more a case of Griffin expressing his frustrations with NASA's progress in returning to the Moon than a reflection of any race with China — after the Chang'e launch, Griffin sent congratulations to his Chinese counterpart. NASA is planning to launch its own Lunar Reconnaissance Orbiter into a low polar orbit late in 2008.

China has made a few overtures towards working in space with other governments, suggesting that it might want to join the International Space Station. The European Space Agency has encouraged such collaboration, for instance allowing China to test its tracking stations for Chang'e by following the 2003 launch of Europe's SMART-1 lunar orbiter.

Any new space race will be mainly in Asia, says Sabathier. In September, Japan launched its SELENE mission, a sophisticated lunar orbiter nicknamed Kaguya, which has reached lunar orbit and released two small satellites. It is checking its instruments and will begin measurements in December. Meanwhile, India is planning to launch a lunar orbiter and impactor in April, called Chandrayaan-1, with potentially a rover to follow.

Eric Hand

See Editorial, page 2.



The Moon may be the target of an Asian space race.

NASA



### COMET BURSTS INTO BRIGHTNESS

Star-spotters beware: that new bright speck is actually a comet.

[www.nature.com/news](http://www.nature.com/news)

C. SCHUR

# India digs deep to build neutrino lab

India's high-energy physicists have been given the go-ahead to build an underground laboratory to study neutrinos.

The plan to investigate these elusive subatomic particles has been in the works since 2001. But the Department of Atomic Energy has now given it the green light and the nation's planning commission has approved its funding, says Naba Mondal, a physicist at the Tata Institute of Fundamental Research in Mumbai and spokesman for the project. It is expected to cost US\$170 million to build and should be online by 2012.

Home for the India-based Neutrino Observatory (INO) will be a man-made cave about 2 kilometres beneath the tallest peak of the Nilgiri Hills, 250 kilometres south of Bangalore. The rock above it will shield the observatory from unwanted cosmic radiation.

Initially, the lab will study neutrinos

produced by cosmic rays in Earth's atmosphere. "A large experiment whose goal is to study atmospheric neutrinos has never been built," says Maury Goodman, a neutrino physicist at Argonne National Laboratory in Illinois. "In that way the INO is unique and important."

The INO will use a 50,000-tonne magnetized iron calorimeter to detect particles called muons that are produced on the rare occasions that neutrinos interact with matter. Because the detector can distinguish between muons with a positive and a negative charge, says Mondal, it should help researchers to work out the respective masses of the three types of neutrino.

John Learned, a physicist at the University of Hawaii at Manoa in Honolulu, says the iron detector is an "excellent choice" that will

also allow the study of neutrino 'oscillations' — when one type of neutrino changes into another as it travels through space.

The INO aims to recapture the pioneering spirit of studies in a gold mine in Kolar,

**"This could bring about a resurgence of the expertise in particle astrophysics in India."**

southern India, which led to the detection of atmospheric neutrinos in 1965. But the facility was abandoned in 1992, and many researchers left for abroad.

Now, India wants to revitalize its neutrino physics community. The INO collaboration currently includes at least 100 scientists from 18 Indian institutions and the University of Hawaii. Learned says the project is likely to bring about "a resurgence of the expertise in particle astrophysics in India".

Project scientists are also in talks with US, Italian and Japanese researchers about collaborating, says Mondal.

K. S. Jayaraman

**<sup>14</sup>C** HIGHEST QUALITY

**50% OFF RADIOCHEMICALS**

LOWEST PRICES

**<sup>3</sup>H**

CATALOG	COMPOUND NAMES (99% PURE)	QTY	PERKINELMER	GE HEALTHCARE	ARC
ART 0207	AMINOBUTYRIC ACID $\gamma$ -[2,3- <sup>3</sup> H(N)]	1 mCi	\$1226	\$1292	\$619
ARC 0290	ARACHIDONIC ACID, [1- <sup>14</sup> C]	50 $\mu$ Ci	\$1090	\$1467	\$549
ART 0196	ARACHIDONIC ACID, [5,6,8,9,11,12,14,15- <sup>3</sup> H(N)]	250 $\mu$ Ci	\$675	\$927	\$339
ARP 0148	CERAMIDE-D-ERYTHRO-1-PHOSPHATE [ <sup>32</sup> P]	10 $\mu$ Ci	ARC EXCLUSIVE!		\$1149
ART 1396	COENZYME A [ <sup>3</sup> H(G)]	10 $\mu$ Ci	ARC EXCLUSIVE!		\$999
ARC 0111	DEOXY-D-GLUCOSE, 2-[1- <sup>14</sup> C]	1 mCi	\$2148		\$1079
ART 0324	FARNESYL PYROPHOSPHATE, [1- <sup>3</sup> H]	250 $\mu$ Ci	\$990	\$1000	\$499
ART 0348	GERANYL GERANYL PYROPHOSPHATE, [1- <sup>3</sup> H]	250 $\mu$ Ci	\$962	\$1488	\$489
ART 0110	GLUCOSAMINE HYDROCHLORIDE, D-[6- <sup>3</sup> H]	5 mCi	\$1356	\$1953	\$679
ART 0116	INOSITOL, MYO-[2- <sup>3</sup> H]	5 mCi	\$2325	\$2954	\$1169
ARC 0126	IDOANTIPYRINE, 4-[N-METHYL- <sup>14</sup> C]	1 mCi	\$2301	\$2400	\$1059
ARC 0541	ISOPENTENYL PYROPHOSPHATE, [1- <sup>14</sup> C]	50 $\mu$ Ci	\$1539	\$1573	\$769
ART 0315	MEVALONOLACTONE, RS-[5- <sup>3</sup> H]	5 mCi	\$2103		\$1059
ART 0129	PALMITIC ACID [9,10- <sup>3</sup> H(N)]	5 mCi	\$240	\$573	\$129
ARC 0772	SPHINGOMYELIN, [METHYL- <sup>14</sup> C]	10 $\mu$ Ci	\$1214	\$1383	\$579
ARP 0144	SPHINGOSINE, D-ERYTHRO-1-PHOSPHATE [ <sup>32</sup> P]	10 $\mu$ Ci	ARC EXCLUSIVE!		\$1149
ART 0778	SPHINGOSINE, D-ERYTHRO [3- <sup>3</sup> H]-1-PHOSPHATE	10 $\mu$ Ci	ARC EXCLUSIVE!		\$1449
ART 0390	STEAROYL [9,10- <sup>3</sup> H] COENZYME A	250 $\mu$ Ci		\$2205	\$1109
ART 0128	UDP-N-ACETYL-D-GLUCOSAMINE, [6- <sup>3</sup> H(N)]	250 $\mu$ Ci	\$1570		\$789



AMERICAN RADIOLABELED CHEMICALS, INC. - 101 ARC DRIVE - SAINT LOUIS, MO 63146 USA  
 TEL: 314-991-4545 - FAX: 314-991-4692 - TOLL FREE: 800-331-6661 - FAX: 800-999-9925  
 E-MAIL: [ARCINC@ARC-INC.COM](mailto:ARCINC@ARC-INC.COM) - WEB: [WWW.ARC-INC.COM](http://WWW.ARC-INC.COM) - REQUEST A COPY OF OUR 2007 CATALOG!

REFERENCE OFFER CODE  
**NAT0711** WHEN ORDERING



## Syngenta shoot-out in Brazil leaves two dead

A gun battle left two dead last week at a Brazilian farm owned by the agricultural biotechnology firm Syngenta. The plantation, in the state of Parana, has been at flashpoint since some 300 activists, protesting against genetically modified crops, first invaded the farm in March 2006 and Syngenta hired guards to protect it.

The shoot-out left one activist and one guard dead, with four more injured on each side, according to news agency reports. A Syngenta spokesman described it as “a quite dramatic and violent confrontation”. The guards’ contract does not authorize them to use weapons.

Although Brazilian law allows some transgenic crops to be grown, the Parana state government has outlawed genetically modified corn and repeatedly attempted to shut the farm down.

## Nuclear-waste programme criticized by scientists

A nuclear-waste reprocessing programme intended to promote a global expansion of atomic energy should be abandoned in favour of basic research, the National Academy of Sciences (NAS) reports this week.

The report underscores many of the criticisms that have been levelled at the Global Nuclear Energy Partnership since it was unveiled by the White House in 2006, namely that it is too risky and expensive for commercial-scale development.

The US Department of Energy (DOE) is seeking \$405 million for the Global Nuclear Energy Partnership for the 2008 fiscal

year, but key Democrats have vowed to put the brakes on it this year. “It seems to the committee that the DOE has given more weight to schedule than to conservative economics and technology,” the NAS panel says in its 29 October report.

## Lack of realism blamed for failed anthrax vaccine

A US government report has concluded that unrealistic expectations were to blame for a small biotechnology company’s inability to fulfil its contract to supply an anthrax vaccine for civilian use.

The 23 October report by the Government Accountability Office found that the Department of Health and Human Services (DHHS) pre-empted development work on the vaccine by awarding the contract to VaxGen in South San Francisco, California, before crucial manufacturing issues had been addressed. Moreover, the Food and Drug Administration’s requirements for data and testing of the vaccine were not known when the contract was approved.

The \$878-million contract was cancelled in December (see *Nature* 445, 12; 2007). As the DHHS prepares to hire another company, the report warns that the department “has not analysed lessons learned from the first contract’s failure and may repeat earlier mistakes”.

## Europe votes to legalize Afghan opium for medicine

The European Parliament has overwhelmingly backed a recommendation for a pilot project looking into the legalization of opium production in Afghanistan to make morphine-related



Flower power: Afghanistan’s opium poppies could become a source of medicines.

medicines, as an alternative to destroying opium poppy crops by chemical spraying.

The vote was 368–49 in favour of the recommendation, which is based on a technical dossier that was produced by the Senlis Council, an international think-tank.

The European Commission currently supports the policy of poppy eradication (see *Nature* 449, 268–269; 2007), which the Senlis Council says has been a disaster. If the Council of the European Union decides to back the European Parliament’s recommendation, the commission will be forced to consider doing the same, says Jorrit Kamminga, the Senlis Council’s head of policy research.

## Genotype data of families with autism are released

Researchers have released a mass of newly generated data providing the most detailed look yet at the patterns of genetic variation in families with autism.

Last week, raw genotype data from 4,250 individuals — including people with autism, their parents and their siblings — were made public by the Autism Consortium based in Boston, Massachusetts, and Johns Hopkins’ McKusick-Nathans Institute of Genetic Medicine in Baltimore, Maryland. Lead researchers Aravinda Chakravarti of Johns Hopkins and Mark Daly of the Autism Consortium, say that they hope to identify key genetic contributions to the disease over the next few months.

In the meantime, says Daly, “we’re releasing raw genotype data so that other qualified researchers can take a look at them even as we’re still beginning our own analysis”. The data will be available through the Autism Genetic Resource Exchange ([www.agre.org](http://www.agre.org)).

## Sale of meteorite meets with stony silence

A piece of one of the world’s most famous meteorites failed to sell at a New York auction on 28 October.

A 13.5 kilogram chunk of the Willamette meteorite (pictured right) was withdrawn after a top bid of \$300,000 fell a long way short of the expected price of \$1 million to \$1.3 million.

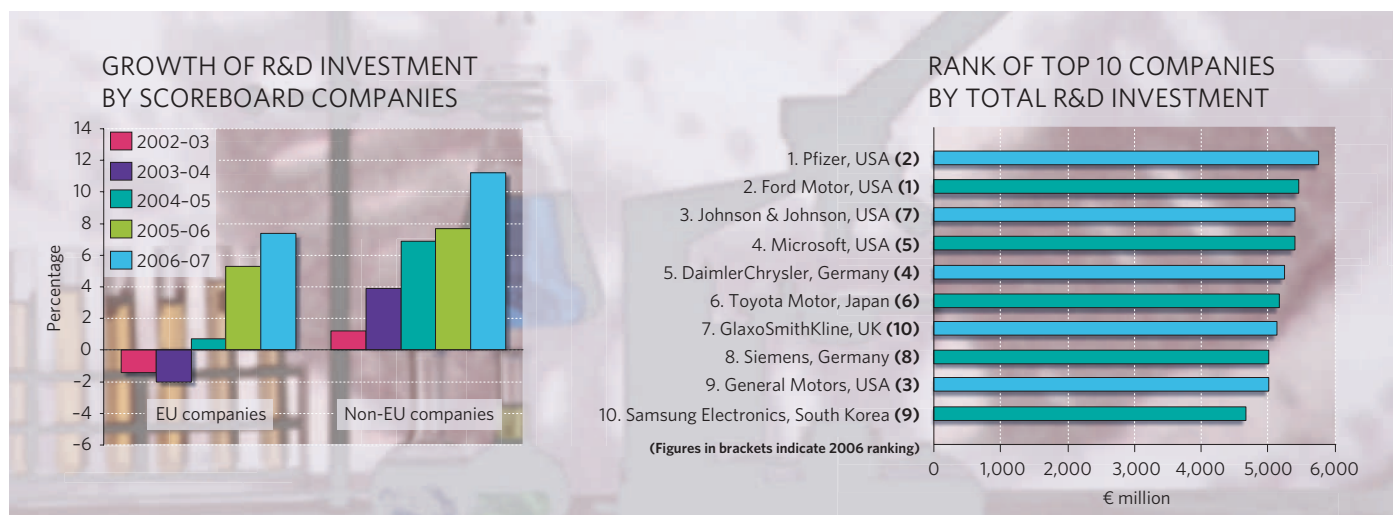
The pockmarked slab of iron and nickel — weighing roughly 14.5 tonnes — from which it was hewn about 10 years ago resides at the American Museum of Natural History in New York. Although ownership of the meteorite has been debated since it was discovered in 1902 (inset), the piece will remain in the Macovich Collection of meteorites, waiting for a buyer.



## BUSINESS

# Mixed scores for European research

A survey shows positive trends for private-sector research and development in the European Union. But as **Andrea Chipman** reports, there's more to the data than meets the eye.



Industrial research and development (R&D) in Europe is surging ahead — but perhaps not quickly enough. Those are the main findings of an exhaustive study undertaken by the European Commission to pin down what big corporate laboratories are up to.

The survey was undertaken after Europe's leaders identified performance in industrial R&D as a crucial factor in keeping European economies competitive.

The 2007 European Union (EU) Industrial R&D Investment Scoreboard, released last month, surveyed the 1,000 biggest R&D spenders in the European Union, and the same number of companies elsewhere.

"We are trying to look at the benefits of R&D," says Héctor Hernández, one of the survey's authors, who is based at the European Commission's Institute for Prospective Technological Studies in Seville, Spain. "We are presenting facts, with a minimum of analysis. It's useful for companies to see where they are — and we are interested in the policy implications," he says.

The exercise is an imperfect one: published data are not always the most accurate guide to the research that actually goes on, and many economists question the correlation between R&D spending and real innovation. But it nonetheless shows up some of the major ways in which global R&D is taking shape.

The scoreboard found, for example, that investment by the top spenders in the European Union rose by 7.4% last year — by far

the highest growth recorded since the exercise began in 2002. But the firms still lagged behind their competitors in the rest of the world, whose R&D investment grew by an astonishing 11.1% in 2006 (see left-hand graph).

The survey also indicated that the pharmaceutical industry is now the largest single industrial sector worldwide in terms of R&D investment, displacing technology hardware and equipment from that slot. Symbolizing that trend, Pfizer replaced the Ford Motor Company this year as the single largest industrial R&D performer (see right-hand graph). The change reflects the drug industry's determination to search even harder for new blockbuster drugs, industry observers say.

Even in the pharmaceutical sector, EU companies lagged behind their counterparts in the rest of the world. They had an R&D intensity — the ratio of investment to net sales — of 12%, compared with 15% elsewhere.

## Obstacle course

"The big difference between the pharmaceutical and other sectors is that we have to reinvent ourselves every 10–15 years," says John Patterson, head of drug development for AstraZeneca in Macclesfield, UK. He points to the industry's challenges from patent expirations and the rising cost of clinical trials as a result of stricter demands from regulatory bodies.

Patterson predicts that the drug industry's position as the top R&D investor will now be consolidated. With around 75% of its main medicines due to go off patent by 2013, the pharmaceutical industry is already spending upwards of US\$60 billion a year and not coming up with enough new products, he says. Patterson's own estimate is that the global industry now spends 16% of sales on R&D — and he reckons this could hit 20% before long. "Either we get more productive or we'll have to spend more," he says.

The scoreboard exercise is part of an EU strategy — first outlined in 2000 and formally enshrined at a meeting of member states' leaders in Barcelona in 2002 — to try to increase R&D in the Euro-

pean Union from just under 2% at that time to 3% of gross domestic product by 2010.

Although this year's scoreboard suggests that some progress has been made, hitting that target now seems to be "practically impossible," says Hernández. But the target still serves a purpose, he contends. "The emphasis of this policy is to raise awareness regarding the importance of boosting research efforts, especially in the private sector."

The scoreboard is compiled with methodology developed at the UK Department of Trade and Industry for its national R&D survey, which has been published annually from 1992. The method limits inclusion to companies with

**"It's not the quantity, but the quality of R&D that is most important."**  
— Nick Von Tunzelmann



audited accounts that are publicly available, and has received few complaints, Hernández says.

The survey includes only companies that itemize R&D expenditure in their financial reports. It also assigns firms to the country of their registered headquarters — even if their R&D operations are located elsewhere. And fluctuations in foreign exchange rates greatly affect companies' relative positions in the rankings. In the case of the depreciation of the US dollar against the euro, for example, the scoreboard underestimates the R&D growth rate of EU companies that have operations in the United States, and overestimates the growth rate of US companies that also operate in the European Union. And as the scoreboard's authors acknowledge, reporting practices and definitions of R&D vary widely between companies and countries.

### Missing links

Specialists also question whether absolute rises in R&D investment actually correlate with competitiveness or innovation. Some argue that the relation between research budgets and financial outcomes is too complicated to be captured accurately in a ranking system such as the scoreboard.

"It's not the quantity, but the quality of R&D that is most important," says Nick Von Tunzelmann of the science policy research unit at the University of Sussex, UK. He admits that such things are hard to measure, but says: "More could be done on the quality side of the scoreboard. The authors could look at how much bang the companies get for their buck."

Von Tunzelmann highlights several factors that make it difficult to draw broad conclusions from the scoreboard. Eastern Asia and the United States tend to dominate the most research-intensive industries, such as information technology, whereas European countries have traditionally been stronger in 'medium-tech' sectors such as automotive production and bulk chemicals.

And although the missing of the 3% target may come as no surprise, the stubborn fact that only 55% of EU R&D was last year undertaken by the private sector — well short of the two-thirds that the European Union had targeted for 2010 — is an even bigger problem, Von Tunzelmann says.

Meanwhile, Von Tunzelmann contends that the role of R&D is changing from an end in itself to a way to develop the capacity to undertake change. Ultimately, he says, companies need to be able to put R&D in its proper context.

"R&D is only a part of the whole innovation effort," von Tunzelmann adds. "On the other hand, it is a fairly strategic part of the whole picture."

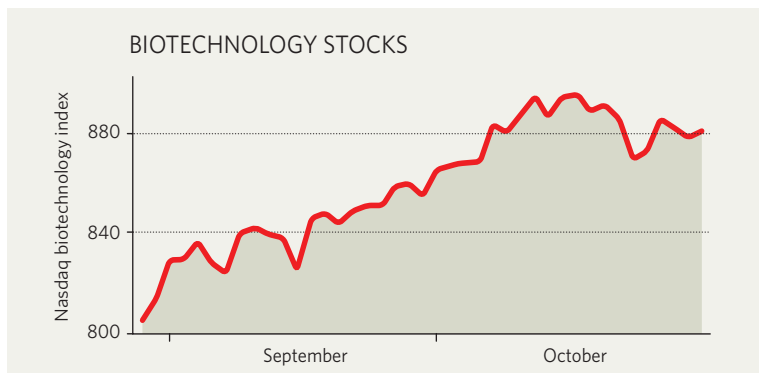
## IN BRIEF

**PATENTS UPHELD** A US jury has ruled decisively for the pharmaceutical firm Amgen in a crucial patent lawsuit, finding that a competing anaemia drug made by Swiss drugmaker Roche infringed several patents belonging to the California-based biotechnology company. After the 23 October ruling, Amgen said that it would seek an injunction to prevent Roche from marketing its drug, Mircera, in the United States. Amgen has held a near-monopoly on the US anaemia market with its drugs Epogen and Aranesp, which together brought in sales of US\$6.6 billion last year. Roche said that it would consider an appeal if, as expected, the court confirms the jury's finding.

**STEEL COUNT** Steel makers have for the first time agreed to set up a system that will collect and report data for greenhouse-gas emissions by the industry on a consistent basis around the world. The International Iron and Steel Institute, based in Brussels, says that it will administer the scheme across the industry, which, it says, accounts for 3–4% of global man-made emissions. However, steel makers have not yet agreed on the extent to which the data for each plant should be made public — some companies have complained that this would require them to divulge confidential information about production levels.

**AUTONOMY LOST** A flagship of British biotechnology disappeared from view on 29 October when Cambridge Antibody Technology (CAT) was merged into MedImmune in Gaithersburg, Maryland. MedImmune is being developed into the biologicals arm of AstraZeneca, the pharmaceutical giant that bought it in April for \$15.6 billion. CAT, which was founded in 1990 and made its name developing antibody therapies, was acquired by AstraZeneca in 2006. CAT's operations will continue, as part of MedImmune.

## MARKET WATCH



This week, Wood Mackenzie, an Edinburgh-based research and consulting firm, reviews recent trends in biotechnology stocks.

The Nasdaq Biotechnology Index has been rising steadily, and is now up 10% from the end of August. This contrasts with a lack-lustre period from May onwards, and has a wide spread, with share prices up at 20 of the 25 most heavily weighted companies in the index.

The share gains reflect speculation that cash-rich pharmaceutical giants are set to buy up biotechnology firms to bolster weak drug-development pipelines and product portfolios threatened by competition from generic drugs.

This perspective was borne out on 12 October when Biogen Idec, the world's fifth biggest biotechnology company, said it was putting itself up for sale. Based in Cambridge, Massachusetts, the firm says that its sale is likely to offer shareholders better value than would be realized by the

company's own strategic growth plans.

Biogen Idec's shares jumped by almost 19% overnight on the announcement, with Pfizer touted as the most likely suitor. Sanofi-Aventis, Novartis, Johnson & Johnson and GlaxoSmithKline are also rumoured to be interested, but Pfizer is the most obvious buyer because drugs that account for about half of its revenue are expected to be facing generic competition by 2011. Of Biogen Idec's five approved drugs, two — Avonex and Tysabri — are treatments for multiple sclerosis, an area of strong strategic interest to Pfizer.

With Biogen Idec currently worth some \$23 billion, any acquisition would be the largest in the drug sector this year, pipping the \$15-billion acquisition of MedImmune by AstraZeneca in April. And investors are betting that it would be just the first in a wave of biotechnology acquisitions by major pharmaceutical companies.



Caught on camera: filmmaker Ruth Berry checks out a tunnel in the Alaskan permafrost.

# Tunnel vision

Is blasting into a river bluff any way to do palaeontology? **Alison Abbott** reports on an unusual expedition into the Alaskan wilderness in search of the bones of polar dinosaurs.

**T**he rabid fox was almost the last straw. In the middle of the Alaskan wilderness, it was tugging at the guy ropes of the mess tent so ferociously that it had torn open its mouth, leaving a trail of bloody spittle across the snow. The gun was in another tent. One person distracted the fox by swinging at it with a frying pan, while another slipped out the back to retrieve the gun. But the ammunition was missing. It took fifteen terrifying minutes to locate it and shoot the fox.

"A comedy of errors" is how Australian palaeontologist Thomas Rich describes the scene. Not funny at all, counters Ruth Berry, the filmmaker who had travelled with Rich to the field site to film his unorthodox plan for getting to dinosaur fossils with the help of some well-placed dynamite. By the time the fox appeared, during the second of three intense expeditions, relations were such that Berry was close to walking away.

Five years ago, the project had looked distinctly more promising. "There are three themes you can always sell at a film market,"

says Berry, who has won assorted awards for her science documentaries. "Hitler, mummies — and dinosaurs." Rich was offering what seemed like a once-in-a-lifetime opportunity: exclusive filming rights to his plan to blast a tunnel into the northern Alaska permafrost above a bed of fossilized dinosaur bones.

At the time, Berry was making a film about the dinosaur bed that Rich, who holds a post at Museum Victoria in Melbourne, had worked for decades on the south coast of Australia with his wife, Patricia Vickers-Rich of Monash University. As Berry researched and shot the film that would be released as *The Terrible Lizards of Oz*, Rich told her how, in the mid-1980s, he had built a tunnel into a fossil-rich cliff face in Australia — and how he wanted to do something similar in Alaska. Such a tunnel, if it worked, would improve access to bones containing unprecedented infor-

mation about polar dinosaurs.

Seventy million years ago the Alaskan landscape was covered with a rich mixture of vegetation including deciduous conifers and ferns, and was home to diverse species of polar dinosaurs. Rich argued that a tunnel could reveal much about the Alaskan dinosaurs and the environment in which they lived<sup>1</sup>.

When Berry, an Australian living in Munich, began shooting her new film, she was convinced of the purity and nobility of science and its processes. She lost this naivety quickly — but her documentary will have a happy ending. This summer she was able to film palaeontologists — although they did not include Rich — working the floor of a sturdy permafrost tunnel and finding what they describe as a treasure of fossilized bones. "The film may have an edge to it I hadn't predicted, though," says Berry. "The tensions in the field were often awful."



Hadrosaur fossils abound in Alaska.

R. BERRY



R. BERRY

Palaeontologists traditionally rely on the natural forces of wind and water to expose fossils, and then work the bones out by hand using pick or brush. On rare occasions they will use earth-moving equipment to get closer to a fossil-rich rock formation. But Rich planned to use dynamite to blast a tunnel above a well-established bone-bed, in order to penetrate deep into the frozen ground to where fossilized bones had not been subjected to seasonal cycles of freezing and thawing.

### New digs

Rich told Berry how difficult it had been to get research-funding agencies to pay for a tunnel. "But somehow I didn't believe him," she says. "I was convinced that the importance of learning about the past history of the Earth would be evident to any research body and that there would be no problem."

R. BERRY

Hawking the idea at film markets was as easy as Berry had thought, and she rounded up around \$800,000 — mainly from Film Finance Corporation Australia and from NOVA, a series produced by WGBH in Boston, but also from smaller investors, including the Australian Broadcasting Corporation and ARTE France. This, she reckoned, would cover the costs of three expeditions to bring film crews and scientists to the remote site.

Their target was the Liscomb bone-bed, located more than 900 kilometres north of the university town of Fairbanks, on the banks of the Colville River on Alaska's North Slope. The site is named after Robert Liscomb, the Shell Oil geologist who discovered fossils there while prospecting in 1961. Liscomb misidentified the bones as belonging to mammoths and died in a rockfall the next year; the fossils languished for years in Shell's archives. They came to light only in 1984, when Shell handed them over to the US Geological Survey, which identified them as 70-million-year-old fossilized dinosaur bones.

The Liscomb bed turned out to contain a whole trove of dinosaur fossils — mostly fragmented skulls and bones belonging to hadrosaurs, plant-eating dinosaurs known as duckbills. Throughout the 1980s and 1990s, the site had been worked by Roland Gangloff, a palaeontologist at the University of Alaska in Fairbanks, who then handed over key work to Anthony Fiorillo, a palaeontologist at the Museum of Nature & Science in Dallas, Texas, who specializes in Alaskan dinosaurs<sup>2,3</sup>.

By spring 2007, Berry was set to go, but Rich had still not secured funding to build the tunnel itself.

So Berry made the daring decision to pay for it with \$167,000 out of her film budget. This meant cutting back on computer animation — a high price to ensure the continuation of her filming project. "But, to be honest, I was seduced by the excitement of the science and the purity of scientific values," she says. "I felt I was doing something good for the world."

NOVA was keen to have an American protagonist in the film as well as an Australian. Berry got in touch with Fiorillo, widely seen as the leading expert on Alaskan dinosaurs, and arranged to film his team at a nearby bone-bed in August 2006, when the group planned to helicopter out a plaster jacket containing a huge horned dinosaur skull. And because the Australian Broadcasting Corporation wanted Rich to be the film's main protagonist, Berry paid for him to fly up and be filmed talking palaeontology with Fiorillo. Only then did she learn that you can bring scientists together, but you can't make them spark. During two days in the field, they didn't speak, in part because Fiorillo was ill. "I hadn't appreciated there would be rivalry," she says. "I'd imagined them talking together excitedly about the great science that might emerge."

Fiorillo and his colleagues clearly felt their site was being invaded by Rich's plan to test the feasibility of building a permafrost tunnel. "It's a bit audacious to go into a well-occupied site solely to see if a technology works," says Fiorillo.

That was a taste of problems to come. In March



Having a blast: Thomas Rich shows off a fossil.

**"The film may have an edge to it I hadn't predicted. The tensions in the field were often awful."**

— Ruth Berry

2007, Berry left on her second expedition to the Colville, planning to bring the heavy mining equipment up on the still-frozen river. When she arrived in Fairbanks with her small film crew and a team from the University of Alaska, she found that the permits were not yet ready — delaying the start for two weeks.

Mike Kunz, the Bureau of Land Management field officer responsible for the Colville area, was also starting to worry that Rich did not understand the difficulties of working in the extreme conditions of the Arctic at that time of year.

"There was consternation about this among the expedition principals," he remembers.

Getting out to the site involves a two-day drive from Fairbanks through increasingly wild locations with names such as Coldfoot, Gobblers Knob and Last Chance. The road, built to follow and service the massive Alaskan oil pipeline, finishes at Deadhorse. From there, it is a four-hour drive southwest to get to a site maintained by the oil company ConocoPhillips, where Kunz arranged for the expedition to stay while a skeleton team travelled the remaining distance to set up the final camp.

### High stakes

Team spirit broke down almost as soon as the full group reached the Colville. One of the stakes Rich had planted in summer a few years earlier to mark where the tunnel should be drilled had disappeared, and the remaining one seemed to be in the wrong place; squabbling ensued. And as temperatures sank as low as  $-40^{\circ}\text{C}$ , and  $20^{\circ}\text{C}$  colder with wind chill, Rich worried team members by failing to adhere to safety instructions, often leaving his tent without the regulation hat and gloves.

Rich says that his experience working in Antarctica had taught him what was acceptable and what was not in extreme weather conditions.

Within ten days, a team of permafrost miners led by Robert Fithian had blasted into the river bluff and created the tunnel. Checking frequently that they were not hitting bones, and altering the angle when that seemed to happen, they carved a volume roughly ten metres deep, three metres wide and three metres high, with insulated thermal walls, a thermal door and a portico jutting out of the cliff face. The tunnel's floor rose towards the back so any water entering in the event of spring flooding could flow out again.

On the last morning, Rich decided to baptize the freshly minted tunnel with a bottle of champagne he'd brought for the occasion. The



others were too busy at the time, so Rich went and opened the bottle alone. Not helpful, as far as the filmmaker was concerned. “We tried to redo the celebration later in the day,” says Berry. “I didn’t film it with the big camera because I thought it would look insincere.” It’s a decision she now regrets.

The final expedition, in August 2007, was to be the climax of the filming. Berry had hoped to capture a glorious moment when the door was opened. She would have liked to film Rich with Kevin May, a palaeontologist from the University of Alaska, locating the bone-bed after some careful digging — and ideally pulling out an extraordinary fossil specimen. But the August trip was no less traumatic than the spring one.

### Floor plan

When the expedition assembled at Deadhorse, it emerged that Rich planned to use power tools to cut half of the tunnel floor into 40-centimetre blocks and transport them to a laboratory in Australia, leaving the other half for May to excavate with his more conventional methods. May was worried that blocks of permafrost might disintegrate when they thawed; Berry was furious at what she saw as a personal betrayal. “I’d invested a lot of time, energy and money in the project,” she says, “and there would be nothing to film if the bed was just carted away.”

Kunz was also concerned. He put Rich’s Bureau of Land Management licence on hold while he checked with experts to ensure that the large air compressor and its supply of diesel Rich planned to have helicoptered to the top of the bluff would not endanger the environment. He cleared that aspect, but put his foot down when it came to cutting the blocks. “These issues had not been addressed in the permit application,” he says, “and when Rich had mentioned it to me a few months earlier I’d informed him it wasn’t going to be acceptable. The blocks would melt into glob and we’d have no idea of the position of bones in relation to each other or to other items in the excavation unit.”

Rich, for his part, wasn’t happy with what he considered “micromanagement”. He says that none of the permit officers had asked for details about the method of excavation and that those on site in Alaska “did not understand the pioneering nature of the project”. He says he thought he could do more work in carefully controlled lab conditions away from the site. “The tunnel is so cold and cramped, it makes everything very slow,” he says.

With the disgruntlement following the team



Researcher Anthony Fiorillo talks fossils on the banks of northern Alaska’s Colville River.

to the bone-bed, the next disappointment was finding the tunnel chock-full of debris and ice — much more than had been anticipated — from spring flooding. May and his team were unable to help dig it out, having had no safety training with a jackhammer. The unskilled workers took as long to clear the tunnel as the miners had taken to dig it from scratch.

With no chance of removing the blocks, Rich left with his team of four a fortnight earlier than planned, even before the team had reached the bone-bed in the tunnel, and transferred his permit to May. By the time the film crew left, Berry did not have her hoped-for ‘eureka’ moment in the can. May then had only six days left to dig with his remaining team, comprising just one museum colleague, his 11-year-old son and Anne Pasch, emeritus professor at the University of Alaska. But he says that was long enough to dig 30 centimetres deep and find around 80 bones in good condition — “much better,” he claims, “than what we would have expected to gather outside, where they would have been subject to successive freeze-thaw cycles.” Although very cold, the tunnel was “a fabulous place to work,” he says. “And although we lost a lot of time this season, next year we’ll be better prepared and we’ll hit the ground running.”

But Fiorillo, who was working on the same bone-bed 75 metres from the tunnel in August, insists that it offers no advantage. “I have been finding intact fossils outside the tunnel without problem,” he says. He adds that the expedition

disturbed his own team: “The noise of the generator they were using to light the tunnel was making us crazy while we were working the bone-bed, not to mention all the helicopter traffic,” Fiorillo and his team were so harried they shifted their camp — “hard to imagine how it could get crowded in the Arctic, isn’t it?” he asks with dark sarcasm — and he says that they couldn’t excavate exactly where he wanted to. He also objected to Rich’s arriving at his technology project with a group of scientists.

But Kunz says that the tunnel worked “spectacularly well”. Berry, although still bristling, also concedes that the tunnel was a success. “In fact, it’s beautiful, but no one can say that it has proved its scientific value yet,” she says. “That will maybe happen next year.”

She says she’ll still have her film, even if the shots from Alaska were not all she had imagined. The back story — about how palaeontologists showed that dinosaurs really existed at the poles and how scientists are now trying to work out how they managed to survive in the cold, dark extremes — is exciting enough for most networks, she says. “In showing the world that dinosaurs really did live at the poles, I’ll have achieved my aim.” ■

**Alison Abbott** is *Nature’s* senior European correspondent.

1. Rich, T. H. *et al. Science* **295**, 979–980 (2002).
2. Fiorillo, A. R. & Gangloff, R. A. *J. Vertebr. Paleontol.* **20**, 675–682 (2000).
3. Gangloff, R. A., Fiorillo, A. R. & Norton, D. W. *J. Paleontol.* **79**, 997–1001 (2005).

See the online version of this story for a short video of Ruth Berry’s film in the permafrost tunnel

► [www.nature.com/uidfinder/10.1038/450018a](http://www.nature.com/uidfinder/10.1038/450018a)

**“Hard to imagine how it could get crowded in the Arctic, isn’t it?”**  
— Anthony Fiorillo





M. LOVETT

When most people look at lobsters, they see dinner. Eve Marder saw a key to the theoretical underpinnings of animal behaviour. **Ishani Ganguli** reports.

Larry Abbott was a theoretical particle physicist at Brandeis University when he first met Eve Marder at an off-campus retreat to celebrate the induction of a new centre for neuroscience. It was the late 1980s and he was just starting to dip his toes into the statistical mechanics of neural networks. But he wasn't sure what to expect from Marder, who had already made a name for herself as an iconoclast in the field. Abbott talked about his mathematical models, and Marder told him he needed to see a 'real' neural network in person, so she invited him to visit her lab.

A few weeks passed back at the Brandeis campus in Waltham, Massachusetts, and Abbott's student needed him about the invite. "I finally listened," he says, and he was blown away.

Along a bench, tiny grape-like clusters of nerve cells pulsed away in Petri dishes. Their activities weren't directly visible, but the rhythmic discharges between nerves ticked away on a chart recorder attached to the cells by electrodes.

"It was so fascinating — what was going on — that I thought, 'I should become a neuroscientist,'" says Abbott. And, with help from Marder, he did. Marder spent hours with

him, meeting daily to answer his questions and later to discuss theories that would define both of their careers. Something for which Abbott, now a long-time collaborator of Marder and a professor of theoretical neuroscience at Columbia University in New York, is grateful: "I was a nobody," he says. "She could have easily given me a quick tour and forgotten about it. But she didn't. She's quite remarkable that way in not giving up on unlikely sorts of people."

For nearly four decades, Marder has taken on all kinds of unlikely sorts in the hope of forging a path of her own design in neuroscience. The

59-year-old has crafted a tremendous body of work on — of all things — the stomachs of lobsters and crabs bought from local fish markets, using the creatures' gastric wiring

as an investigational model to launch several fields of inquiry and dozens of careers. Her work has explained how even simple neural networks can create diverse functional repertoires, and how the homeostatic mechanisms within allow these networks to produce reliable behaviour patterns despite the constant turnover of nerve-cell components.

Marder first found herself at Brandeis as a

teenager in autumn 1965. Any scientific aspirations she had before going to college withered in the political heat of the time, she says. Demonstrations and voter-registration drives led her to dream of becoming a civil-rights lawyer. She declared a major in politics.

But in her junior year, Marder followed a room-mate into a course on abnormal psychology that changed everything for her. These were the days when the blame for schizophrenia and many other psychological ills was placed squarely on parents. Marder was intrigued by the idea, then heretical, that schizophrenia had a biochemical cause and that dysregulation of neurotransmitters — which relay electrochemical signals from nerve to nerve — could vastly influence a person's grip on reality. So she holed herself up in the Brandeis library and read every book she could find on how the brain seemed to use the nerve signal-dampening inhibitory neurotransmitter, gamma-aminobutyric acid (GABA) to keep itself in check. This first introduction to neurotransmitters sparked a lifelong curiosity about nerve-cell communication. She switched her major to biology and never looked back.

In 1969, Marder started graduate school at the University of California, San Diego. It was there, under Allen Selverston, that she

**"It was so fascinating that I thought I should become a neuroscientist."**

— Larry Abbott

first encountered the lobster stomatogastric-ganglion (STG) system. This neural network comprises just 30 neurons that control the gut muscles of lobsters and other crustaceans — allowing them to grind food using gastric ‘teeth’ and then pass it down their digestive tract. It has become the best-studied example of a central-pattern generator, the same type of rhythmic neural circuit that controls breathing and chewing in humans. Excised and secured on a Petri dish, this network will fire rhythmically to the beat of its own drum for hours, without any outside input. The STG’s large neurons, easy to identify and record from the preparation, provide a ready glimpse into the tiny but powerful circuit.

### The big picture

Whereas Marder says that most people in neurobiology were studying single neurotransmitters such as serotonin and GABA in isolation, she had broader questions in mind. As many as eight neurotransmitters had been identified by then, and she wanted to know why so many were required in the brain.

As she was discovering the system that would define her research career, Marder was also working to define her values as a scientist. As a graduate student, she watched fellow neuroscientists scramble to purify the nicotinic acetylcholine receptor — a critical part of nerve and neuromuscular signalling — and quickly learned something about herself. “[With] a consensus problem that everybody thinks is interesting, you have a built-in audience and built-in peer group approval. But I realized I would never choose to work that way,” she says. Her own vision would matter little, Marder says, in an “enter-prise where the goal was so well defined that it was going to get done no matter what”.

So she went after the circuitry of the STG system. Glutamate had already been established as the neurotransmitter at

play at many of the synapses — the connections between nerve cells. With painstaking measurements of electrical activity in the 30 cells of the ganglion, and by searching for biochemical clues, she discovered that it wasn’t alone. Acetylcholine — the neurotransmitter of choice in the nerve to muscle interfaces of vertebrates — was also at play in half of the neuromuscular junctions in this arthropod system. She found that like glutamate, acetylcholine could excite or inhibit depending on context, adding to a growing body of evidence that a neurotransmitter’s effect could vary based on its targets<sup>1</sup>.

The findings earned her her first publication — in *Nature* — and a PhD. Marder continued her pursuit of neurotransmitters over four years of postdoctoral training at the University of Oregon in Eugene and the École Normale Supérieure in Paris, France, where she honed her biochemistry, electrophysiology and biophysics techniques. In 1978, Marder turned down a tenure track position at Cornell University in Ithaca, New York, to return to her alma mater as an assistant professor, and has not left since.

Neuroscientists had always treated networks as electrical systems with fixed connectivity. But an offhand observation Marder had made as a graduate student led her to question this dogma. In the process of figuring out which neurotransmitters the ganglion cells use to communicate, Marder had dumped candidate after candidate onto the system. She had noticed at the time that many of these molecules had an effect on the output of her tried and true lobster preparation, even those that weren’t acting as neurotransmitters. But she hadn’t known what to make of this at the time. Newly situated at Brandeis, she began to investigate with her budding lab team.

What she soon began to realize was that these substances — acting on a time course of hundreds of milliseconds to hours, a glacial pace in the firestorm of STG signalling — were ‘neuromodulators’. This was heresy. “People working in vertebrate systems

still thought that vertebrate neurons ... were very simple on-off figures.” But modulators meant that the systems didn’t have to be hard wired. Neurons can release one or several neuromodulators — some of which are also neurotransmitters, such as serotonin, and others, such as the peptide proctolin. Like hormones, they bind to receptors on other neurons, triggering long-lasting changes in how the neurons respond to the fast-acting neurotransmitters that allow cells to communicate. “They can do all sorts of interesting things,” she says, such as alter

the intrinsic excitability of cells, the amount of neurotransmitter released each time a neuron ‘fires’, or the firing patterns as a whole. She began to parse out which neuron types have which receptors for these functions<sup>2</sup>.

New antibodies against potential modulators eased the process of searching intact tissue for these modulators through the late 1980s and early 90s. “Every year or two would add one or two more,” says Marder. Mass spectroscopy accelerated the search in the late 1990s. Within the next 10 years, she says, the full cast of 50-odd characters in the neuromodulation story should be assembled — by her lab and others. At that point, she says, they will get a clearer picture of how all the parts interact.

### From lobsters to humans

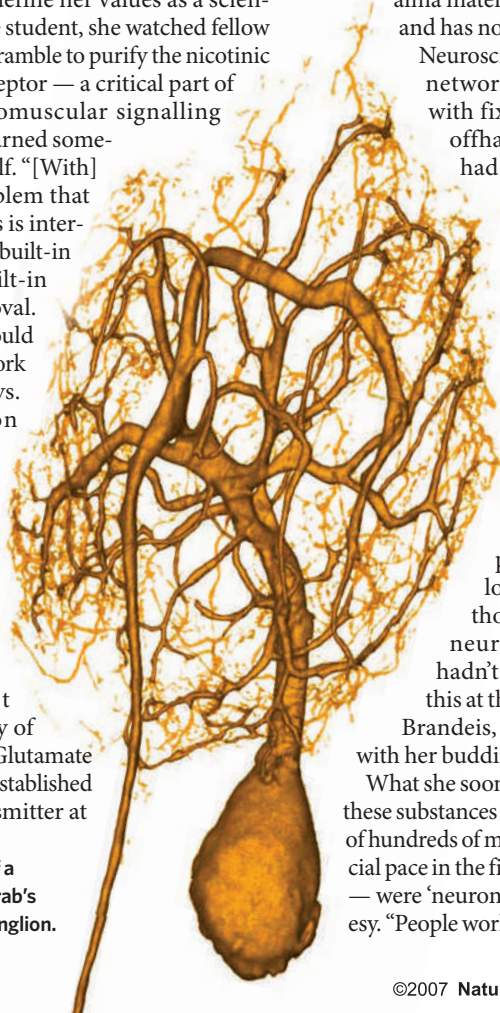
The applicability of her work on crustaceans to human neuroscience is a point that animates her. Neuromodulators have been found in all animal brains. “If you don’t think they’re doing the same thing, you’re just silly,” she says. “As far as I’m concerned, evolution never throws away a good mechanism.” Although experts were not open to the neuromodulators’ snail-paced mode of communication between neurons, Marder considered it a logical step. “It was so obvious to me that this is the way it was working that I was always surprised when people didn’t think it worked that way.”

Abbott, at least, had no reason to doubt her, and months after they met, their collaboration spawned the birth of the ‘dynamic clamp’: a neurophysiological method that allows researchers to simulate finely tuned neural networks using computer-manipulated nerve cells. Scientists can adjust different parameters, tweaking conductance through ion channels triggered by ligands or voltage changes, and see the effects on the circuit in real time<sup>3</sup>. The clamp is now used worldwide in diverse applications such as simulating heart muscle cells.

At the same time, her collaboration with Abbott led her to ask new questions about the

**“Marder has an amazing ability to come up with ways to get at big questions that are relevant to neuroscience.”**

— Adam Taylor



Impression of a neuron from a crab's stomatogastric ganglion.



neurons she had studied for decades. The neurons were resourceful in modifying their patterns, but how did they manage to keep their basic performance intact over time? Marder had been working with lab members to build a model of the lateral pyloric motor neuron, an important controller of the stomatogastric system. "I found it frustrating beyond belief," Marder recalls. Coordinating the individual current contributions of ion channels in each membrane to predict the system's overall behaviour proved unwieldy. "I said if the model is so sensitive to variations of each type that it takes a smart postdoc months and months to play with, how does the neuron get it right?"

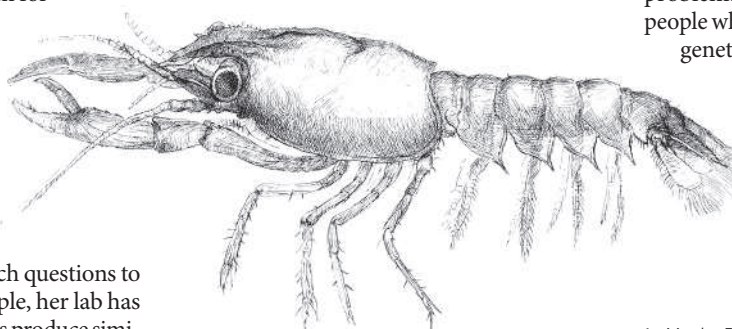
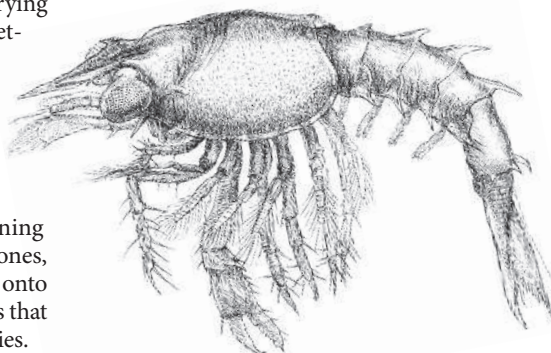
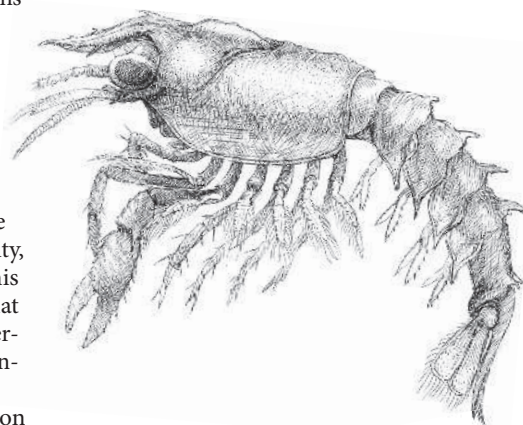
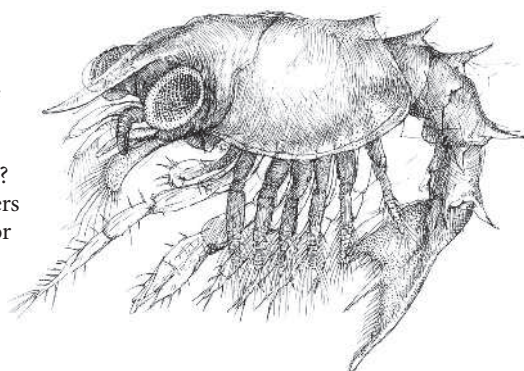
"There have got to be some simple rules that let the neuron self-tune even though ion channels are turning over continuously over time," she remembers venturing to Abbott. It turns out there were. The independent variable was the output activity, not the number of channels mediating this activity. Marder and Abbott discovered that neurons were remarkably resilient when perturbed, changing their conductances to maintain a given output<sup>4</sup>.

What are these rules by which the neuron gets its activity right? This question of homeostasis continues to drive her work. She is trying to figure out how cells maintain stable network performance over long periods of time despite the ongoing replacement of ions and ion channels.

Today, as when she started, Marder says that about a dozen labs are working on the stomatogastric ganglion system. The niche field suits Marder's style: turning unexpected questions into mainstream ones, then passing the pursuit of their answers onto researchers working on model organisms that can best handle them, such as mice and flies.

"One of the dangers of working on a prep such as the STG is if you're not careful, you can fall into studying the system for the system's sake," says Adam Taylor, a postdoctoral fellow in Marder's lab. "She has an amazing ability to come up with ways to get at big questions that are relevant to neuroscience writ large within the stomatogastric ganglion of crustaceans."

And there are many more such questions to answer, says Marder. For example, her lab has shown that different mechanisms produce similar output patterns, but to what extent does this concept extend to more complex systems? "Your respiratory system keeps you breathing, mine keeps me breathing. How different are they



**Larval stages of the Maine lobster (*Homarus americanus*), the subject of much of Eve Marder's work on crustacean stomatogastric ganglia.**

really?" she asks. She is also trying to understand how this built-in variability fits into the preservation of circuits as animals grow and develop. The neurons of adult lobsters are much bigger than those of juveniles, yet she has measured the same motor patterns in both generations, suggesting that the animal retunes properties such as cell-membrane resistance or the distance between nerve cells to produce the same effect.

### Do unto others

Taylor says Marder leads by example. "It's quite a trick," he says, how she "manages both to make you feel like you should work harder and not make you feel depressed about your boss being a slave-driver". Such skills may come in handy in November when she adds president of the Society for Neuroscience to her extracurricular activities, which already include editing the *Journal of Neurophysiology* and service on several advisory committees and review boards.

Marder has a habit of downplaying her accomplishments — which include induction into the National Academy of Sciences earlier this year, and winning the 2005 Ralph W. Gerard Prize from the Society for Neuroscience. During an interview at her home in down-town Boston, her husband — Arthur Wingfield, also a professor of neuroscience at Brandeis — gently chides her for her reticence as he lists her honours and responsibilities.

But her spotlight-shyness is part of a desire to be so far out on the leading edge that no one notices her. At least not right away. It's a challenge she revels in. "If you work on a non-consensus problem, you have the additional burden of having to do something that's interesting enough, novel enough or articulate enough to change the way people think," Marder says. Certainly her work has articulated enough questions to keep others busy. "There are parts of these problems that will be much better solved by people who can work on mice and flies and real genetic organisms."

For her own part, though, Marder's model loyalty is firm. It even extends beyond the bench. "I don't like to eat lobsters anymore," she says, "because I find I just feel bad for them."

**Ishani Ganguli is a freelance writer in Boston, Massachusetts.**

1. Marder, E. *Nature* **251**, 730–731 (1974).
2. Marder, E. & Eisen, J. S. *J. Neurophysiol.* **51**, 1345–1361 (1984).
3. Sharp, A. A., O'Neil, M. B., Abbott, L. F., & Marder, E. *J. Neurophysiol.* **69**, 992–995 (1993).
4. LeMasson, G., Marder, E. & Abbott, L. F. *Science* **259**, 1915–1917 (1993).



# A COUNTRY WITHOUT ALMS

The Japanese make few charitable donations. David Cyranoski meets a patient advocate and scientist working to change a cultural reticence about giving.

**M**akoto Ohama has two major goals in life. One is to feed himself, using his own hands — the 62-year-old was paralysed from the neck down in a rugby tackle 30 years ago and knows that this goal would require a leap in regenerative therapy. The second, as president and chief executive of the Japan Spinal Cord Foundation, is to accelerate the research that might one day lead to such treatment, for others if not for himself. But, in Japan, this more modest goal seems equally remote.

Despite being Japan's leading non-profit organization devoted to spinal-cord injury, Ohama's foundation can barely afford to pay for administrative and community-building activities. It has no money for scientists. In Japan, there is a trifling amount of philanthropic money for foundations to draw on. In 2005, the twenty largest grant-giving foundations in Japan spent only one-thirtieth of what their counterparts in the United States did, even though Japan has one-third of the United States' economic might. And most philanthropic funds in Japan are tied to industry, creating potential conflicts of interest. Unlike foundations in the West, where private donations "just pour in", Ohama says that in Japan, "the will to give is weak".

This is particularly bad news for researchers who work in unpopular fields, including those stymied by political debate, such as stem-cell research. Japan's government does fund research at levels comparable to the governments of most other developed countries — 0.67% of gross domestic product, compared with 0.83% by the

United States. But foundations play a crucial part in filling the holes left by the government and in supporting young researchers, says Hideyuki Yuze, deputy director of the Tokyo-based Japan Foundation Centre. They also fund research on rare diseases in the developing world that pharmaceutical companies may ignore. "They can give long-term support and balance political inconsistencies," says Shin-ichi Nishikawa, a stem-cell expert at the Center for Developmental Biology in Kobe.

Like Ohama, Nishikawa is hoping to boost philanthropic support for science in Japan. But in their quest, the two campaigners will have to take on economic realities, government regulations, tax policies and most intractably, a culture in which individuals, rich or not, do not generally donate.

The richest Americans are high-profile donors. Bill Gates donated some \$11 billion between 1995 and 2005, much of which went to the Bill & Melinda Gates Foundation in Seattle, Washington. Warren Buffett last year pledged \$43.5 billion — the largest philanthropic gift ever — to several organizations, including the Gates foundation (see [www.nature.com/news/specials/philanthropy/index.html](http://www.nature.com/news/specials/philanthropy/index.html)).

Japan's big givers are notable by their absence. "We just don't have the extraordinarily rich," says Yuze. Japan's richest person, Masayoshi Son, a self-made communications mogul worth \$5.8 billion, does not give to charity, except in times of disaster such as earthquakes. He ranks 129th

on the Forbes 2007 list of billionaires. A spokesperson for Japan's second richest person, real-estate mogul Akira Mori (\$5.7 billion), refused to say whether Mori or the Mori Trust, of which he is president, make charitable donations.

"Rich people are not respected in Japan, so they hide their money," says Yoko Takahashi, director of the Japan Philanthropic Association in Tokyo. "Many of the people who give do so anonymously," she says. "I'd like to see rich people applauded for doing things."

But the lack of charitable giving extends to citizens further down the economic ladder. Fund-raising efforts aimed at the average citizen generally fall flat. It took 3 years of monthly campaigns for Ohama to raise

the ¥3 million (US\$24,000) necessary to establish his non-profit organization in 1997. His foundation now raises an average of ¥21 million per year, but half of that comes from grants and subsidies. His biggest fund-raising event was an Internet campaign that brought in just ¥700,000. Other organizations, such as the Kidney Foundation in Tokyo, do not bother with fund-raising because they lose money on it.

## Big bucks

By contrast, the Miami Project to Cure Paralysis in Florida can bring in millions of dollars from a single fund-raising event, and 85% of its proceeds go to science. In Canada, which has only 40% as many patients with spinal-cord

**"Rich people are not respected in Japan, so they hide their money."**

— Yoko Takahashi

K. SHIROBAYASHI/GETTY IMAGES



injuries as Japan, the Vancouver-based Rick Hansen Foundation's yearly walking and wheelchair-riding Wheels in Motion campaigns bring in an average of Can\$1.5 million a year — eight times more than Ohama's foundation raises in an average year. Half of that goes to research on spinal-cord injuries.

Japan's economic policies discourage philanthropic activity. A low-interest-rate policy, in effect since Japan's economic bubble burst in the 1990s, has cut into income on endowments, and many foundations have had to make drastic cuts or terminate their grant programmes. "They cannot afford to be adventurous," says Yuze.

And strict regulations mean that very few of the country's 32,000 non-profit organizations are eligible to receive tax-free donations. The Japan Spinal Cord Foundation is not one. "The government doesn't trust non-profits," says Motoyuki Watanabe, the foundation's executive director. The regulations were made to discourage money laundering, but those in the foundation business say that it's time for a change.

Nishikawa was part of the lobbying effort that won a small change in the regulation of non-profit organizations, increasing the number receiving tax-free donations from 32 in 2003 to 71 now. He would like to see further changes that would allow several organizations to band together in their fund-raising and lobbying activities.

But the tax situation is not the only reason the average Japanese person is less likely to give than the average American. Unlike in the United States, the national health-care system and generous overseas development aid in Japan may convince people that the government is taking care of charity both at home and abroad. "People think the government is going to do everything for them, but we can't always rely on the government," says Takahashi.

This belief in collective, rather than individual, action might explain why most Japanese foundations are either created by large companies or

rely on them for their funding. Pharmaceutical giant Takeda has one of the biggest — the Takeda Science Foundation — which gave out ¥1 billion in biomedical research grants in 2005. Other companies, such as Mitsubishi and Toyota, have foundations that give less than half that amount. These companies are also members of the Nippon Keidanren's 1% Club — a 'gentleman's agreement' calling on members to give 1% of their profits to society. (Keidanren is an industry association with 1,600 members, including many of Japan's largest companies.) Surveys by the club show that members who donate tend to give 1.3% of their profits to support some kind of social activity, but this does not generally mean donations to foundations.

### Keeping clean

Most 'independent' foundations with enough money to distribute research grants still rely on industry for most of their support. The Kidney Foundation, for example, gets 90% of its money from drug companies. But this tends to restrict funding to research that has a quick pay-off. "If we had more money from individuals, we could look with a longer view," says Mami Honda, the foundation's director. Foundations tied to industry also sacrifice their independence. "You have a hard time decontaminating yourself," says Robert Goldstein, chief scientific officer at the New York chapter of the Juvenile Diabetes Research Foundation. "You lose your clout."

Indeed, in September, the Tokyo-based Japan Heart Foundation was embroiled in a scandal when it was discovered that it had provided ¥200 million for a clinical trial of a beta-blocker for treating heart failure while receiving ¥300 million from the company that made the drug. The health ministry warned the



**Makoto Ohama struggles to raise funds for his foundation for spinal-cord research.**

foundation that such "fund-raising was inappropriate for a charitable organization". The researcher running the trial also broke ministry policy by combining the foundation money with ¥119 million in grants — funds that are not supposed to be combined with any other grants. The trial was completed in 2006,

but the health ministry pulled its funding for the researcher for the following two years.

Remaining independent is key for an organization such as Ohama's because he wants to lobby the government for more funding for spinal-cord-injury research. He also wants to push for looser restrictions on embryonic stem cells: Japan ostensibly has lax policies on stem-cell research but regulatory hurdles stymie the work. The nation has only three established stem-cell lines, but researchers need separate facilities even to work with these (see *Nature* **438**, 263; 2005). They also need to obtain approval every time they change the research plan, or add new people to the team. "There are too many restrictions," says Ohama. "Doctors here can't say that, but as patients we can." But without money his lobbying power is weak.

Nishikawa says that part of the responsibility lies with physicians. "Doctors and researchers do not regard patients as an important group to work with. There's no solidarity with patients," he says. "This should be changed by the doctors." He has given talks at various academic society meetings encouraging scientists to reach out to patient groups. But he faces an uphill battle. Japanese patients are reluctant to come forward with their problems because they fear drawing attention to themselves or their families, especially in the case of genetic disease.

Many Japanese people are just not aware of the issues, says Honda. There are no sports heroes or movie stars willing to use their celebrity to raise awareness in the way that Michael J. Fox did for Parkinson's disease and the Reagan family did for Alzheimer's disease. Honda says that she has asked famous people with kidney problems for help in promoting the foundation's activities, but all have refused.

Nishikawa relates the story of Hisao Niura, the former baseball pitcher for the popular Yomiuri Giants in Tokyo, who initially hid his diabetes. When others noticed a combination of needles and weight loss, rumours started to spread that he was a drug user. "I can't think of any other famous Japanese with disease," says Nishikawa. For now, Ohama's foundation uses Christopher Reeve as its symbol.

**David Cyranoski is *Nature's* Asia-Pacific correspondent.**

JAPAN SPINAL CORD FOUNDATION

K. KASAHARA/AP



**Japan's richest person, Masayoshi Son (left), and America's richest, Melinda and Bill Gates, have different attitudes about donating money to charitable foundations.**

J. NGWENYA/REUTERS



## Sea-ice decline due to more than warming alone

**SIR** — The dramatic loss of sea-ice cover over the Arctic this summer was widely reported, for example in your News story 'Arctic melt opens Northwest passage' (*Nature* **449**, 267; 2007), and frequently attributed to global warming. Although the gradual decline in sea-ice extent during the past four decades is in line with that expected from global warming, it is very unlikely that the loss of sea-ice cover this year is explicable solely in terms of temperature change.

Changing wind patterns are an important influence on the distribution of sea ice. Throughout summer 2007, exceptional pressure and wind patterns persisted over the Arctic Ocean. The observed migration of ice cover, from the Siberian and Beaufort seas northwards and eastwards into the Arctic Basin, was in line with the expected response to the anomalous winds. These Arctic wind anomalies were part of a global-scale pattern of highly unusual circulation this summer, the causes of which are as yet unclear.

The growing La Niña in the East Pacific undoubtedly had a major influence globally, and there is some evidence from past events that La Niña predisposes the circulation towards the type of exceptional patterns seen this summer.

**Julia Slingo, Rowan Sutton**

National Centre for Atmospheric Science, Walker Institute for Climate System Research, University of Reading, Earley Gate, Reading RG6 6BB, UK

## Ethics reviews can be centralized without delays

**SIR** — The problems inherent in continuing the current system of local ethics reviews for multicentre trials are appropriately highlighted in your News Feature 'Trial and error' (*Nature* **448**, 530–532; 2007). However, you state that the Central Institutional Review Board of the US National Cancer Institute (NCI) adds bureaucracy by its efforts to centralize review. We disagree.

A unique feature of our model is that both central and local boards can be involved; the extent of the local review remains the prerogative of the local board, and their task is facilitated by access to the central board's review. This model has had widespread acceptance: 66% of NCI-designated cancer centres have now joined. The institute set up a central board for adult and paediatric cancer treatment trials because of the redundancy that occurs when each site's local institutional review board performs a separate ethics review. For example, 50 to 200 sites typically participate in NCI-sponsored phase III trials.

We do not agree that this centralization

delays the time it takes to activate a study. The institute's Cancer Therapy Evaluation Program reveals that, for more than two-thirds of the protocols reviewed by the central board, activation is not delayed, as numerous other activities continue in parallel, such as creation of case report forms, finalization of contracts with pharmaceutical partners and drug shipment plans.

We believe that centralized ethics review, with local context supplied by the local board, is the preferred way to conduct review of multisite protocols. The NCI is currently supporting a formal cost analysis by an independent third party. In view of the high local costs cited in your News story, such data should compel administrators of local boards to reconsider their position on centralized review.

**Jeffrey Abrams, Jacquelyn Goldberg**

Central Institutional Review Board Initiative, National Cancer Institute, 6130 Executive Boulevard, Bethesda, Maryland 20892, USA

## Missing the mark on biomedical research

**SIR** — Brian Martinson in his Commentary 'Universities and the money fix' (*Nature* **449**, 141–142; 2007) suggests that short-sighted greed for federal grants motivates academic leadership to sacrifice the future interests of biomedical research. On the contrary, such grants do not cover the full costs of the research projects they support.

The annual shortfall of billions of dollars (see C. A. Goldman and T. Williams *Paying for University Research Facilities and Administration* Rand, Santa Monica; 2000) must be offset from the limited sources of discretionary funds available to universities: tuition fees, state funding, philanthropy, technology transfer or revenues derived from the clinical practice of the faculty.

Of these, philanthropy is serendipitous and generally attracts investment in new initiatives rather than in sustaining existing programmes. The pressures on tuition costs and clinical-practice margins have been widely publicized. Technology transfer only rarely brings significant, sustainable revenues.

If academic institutions merely sought to maximize short-term revenues, they would not make the substantial long-term commitments of capital to physical plant, state-of-the-art technologies and skilled personnel that they do.

We believe that the expansion of research capacity is being driven instead by the explosive pace of advances in the biomedical sciences and in our understanding of the processes of health and disease, as well as by mounting dependence on multidisciplinary research teams and increasingly sophisticated new technologies.

These investments are inherently risky in their anticipation of future returns from federal and other sponsored research. Accordingly, the decision-making behind them is deliberate and calculated: university governing boards, bound by their fiduciary obligations, are inherently conservative.

The distress caused by National Institutes of Health budgets that have steadily declined in purchasing power since 2003 is being widely and acutely felt. But increases in these budgets are unlikely; simply sustaining the enterprise to keep pace with inflation — which Martinson terms "prudent" — has been difficult.

The biomedical research enterprise does need to adapt to fiscal realities, but restructuring the historic 'business model' for academic research, including the composition of its workforce, will be very difficult. At minimum, such restructuring must be guided by sensitivity to the institutions' missions, recognition of the intense sociopolitical expectations of them and accurate perception of what is necessary for their academic and economic survival.

**David Korn, Stephen J. Heinig**

Association of American Medical Colleges, 2450 N Street, Washington DC 200037, USA

## Words of wisdom worth reading aloud

**SIR** — I congratulate Erez Lieberman and colleagues on their entertaining, enthralling and, above all, well-written Letter 'Quantifying the evolutionary dynamics of language' (*Nature* **449**, 713–716; 2007). *Nature* articles outside my sphere of interest are of interest to me as far as the abstract; after that I am rapidly lost in a forest of jargon, technicalities and poorly articulated background information.

I had little prior interest in linguistic evolution, yet Lieberman and colleagues captured my attention to the end of their Letter. It was so well written that I was compelled to read out the cleverest snippets to my office-mates.

This is a scientific paper that has truly exemplified the aim of *Nature* to provide a forum for work with interdisciplinary appeal. I congratulate and thank both the authors and the selecting editors.

**Vanessa S. Solomon**

Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne, Victoria, Australia

**Contributions to Correspondence may be submitted to [correspondence@nature.com](mailto:correspondence@nature.com). They should be no longer than 300 words, and ideally shorter. They should be signed by no more than three authors; preferably by one. Published contributions are edited.**

## BOOKS &amp; ARTS



From Steven Pinker's latest: naming a child is an opportunity to leave our linguistic mark.

## Poetry in motion

Is language the key to thought? Neuroscience suggests it is probably more complicated than that.

### The Stuff of Thought: Language as a Window into Human Nature

by Steven Pinker

Viking: 2007. 512 pp. \$29.95

#### Patricia Churchland

The conventional wisdom in psycholinguistics in the 1980s was that thought is like external language in all important respects. Each of us, the argument went, comes genetically equipped with a 'language of thought' that is reflected in the structure and organization of speech. Thought is not remotely similar to perception or imagery, or to the exercise of motor skills. The basic rules governing human thought and language were believed to be largely unique and substantially innate, the result of genetic novelty. Understand language, and — the psycholinguists used to say — you understand thought.

In this line of reasoning, languages relate to the world as follows: names denote, as *Henry VIII* denotes Henry VIII; type terms, such as *planet*, refer to the set of all actual planets. Reference, singular or general, is supposedly fixed when a single person first coins a word — for example, *planet*, while pointing to Jupiter. The proper scope of that term is then said to include all things that 'have the same nature' as Jupiter, where the relevant sameness relation is said to be fixed by physical factors (probably unknown). Were it not so, the story goes, I would not mean what you mean by *planet*, so communication would founder. Fortunately,

says this argument, the ancient Greeks did mean exactly what I mean by *planet*, owing to one having cleverly dubbed Jupiter a planet. Unfortunately for this theory, the Greeks also called the Sun, but not the Earth, a planet.

This approach to word meaning is about as applicable to real meaning as 'Dungeons and Dragons' is to real life. Aptly ridiculed by critics as 'font-change semantics', the theory still has its disciples. Including Steven Pinker.

Indeed, it is essentially font-change semantics that Pinker defends and deploys in his latest engaging doorstop, *The Stuff of Thought*. He has revised a few features, but the core ideas — innateness *ad libitum*, and the quest for the nature of thought in the analysis of language — are intact. Like his earlier books *The Language Instinct* and *The Blank Slate*, *The Stuff of Thought* has very little to do with the stuff with which we think — namely, neurons.

In leaving neurons out of the story, Pinker is not alone. Jerry Fodor elevates ignorance of neuroscience to a methodological virtue, proclaiming, "If you want to understand the mind, study the mind ... not the brain, and certainly not the genes". His metaphor, embraced by some psychologists and philosophers, says that the brain is merely the hardware that happens to implement the cognitive software. Neurons and their connectivity are as irrelevant to understanding the nature of mental function as a computer's transistor configurations are to my using Powerpoint.

Advances in neuroscience and genetics during the past 30 years have put such thinking

on the defensive. For one thing, extravagant claims about human uniqueness must deal with the discovery that humans have only about 28,000 genes, and differ from mice in just 300 or so. Additional constraints emerge with the discovery that human brains are stunningly similar to other mammalian brains — in components, connectivity, development, biochemistry and physiology. Topographic maps in the neocortex, cerebellum, spinal cord and subcortical structures are standard for representing and computing with neurons. As such, they suggest constraints relevant to semantics and reasoning.

Maps that represent which parts of the body are receiving what kind of stimuli are probably crucial to the very nature of our self-representation and in what we mean by self. The pathways connecting sensory maps to those representing motor preparation are likely to be important for reasoning what to do next. Meanings, as W. V. O. Quine realized some 50 years ago, fundamentally relate to the world not piecemeal (*planet* means planet) but through connected representational networks in the brain that, with varying accuracy, map as a whole on to reality. These are the maps that get us around the physical and social worlds.

And getting us around is the basic evolutionary rationale of nervous systems. Unlike plants that must take what comes, animals are movers. More sophisticated behaviour emerged with improved capacities to plan, predict and draw on past experience, which improved chances of surviving and reproducing.



This observation motivated neuroscientist Rodolfo Llinás, in his 2002 book *I of the Vortex*, to propose that, at bottom, thinking is the evolutionary internalization of movement. He meant that thinking is the generation in the brain of images of a future action, and its consequences. And generating these images depends on flexibility in categorizing the current problem as an instance of one kind of event rather than another, which, in turn, depends on memory for past experience. Fundamentally, thinking is neural activity in the service of behaviour (for example, should I flee or fight? Is this attacker weak or strong?). This almost certainly shapes thinking that seems detached from motor preparation (such as, where did Earth come from?).

As is so often the case in biology, discover-

ing structure is crucial in coming to understand function — as in William Harvey's seventeenth-century revelation that hearts are actually pumps, not biological cauldrons for concocting animal spirits. To figure out how brains actually think and what reasoning really amounts to, we need to focus on understanding their many levels of organization, from neurons to large-scale systems to behaviour. If thinking is rooted in internalized movement, it may be more akin to a skill than to a syllogism. Language may not be the "stuff of thought" after all.

Patricia S. Churchland is professor of philosophy at the University of California San Diego, La Jolla, California 92093, USA. She is the author of *Brain-Wise: Studies in Neurophilosophy*.

electrical cells and synaptic spaces", and of forgetting that "this isn't how we experience the world". Some scientists maybe, but it hardly describes psychologists. The coda begins with the "incommensurate languages" of C. P. Snow's *The Two Cultures*, and then attacks Richard Dawkins, Brian Greene, Steven Pinker and E. O. Wilson and the "third culture" for being antagonistic towards everything non-scientific and lacking a dialogue of equals.

By contrast, Ian McEwan's *Saturday* is invoked as a rare cultural commodity that possibly symbolizes the birth of a new fourth culture, in which the humanities engage sincerely with the sciences. Lehrer argues that every humanist should read *Nature* and that scientists should recognize truths other than their own, with art "a necessary counterbalance to the glories and excesses of scientific reductionism, especially as they are applied to human experience". It's a grand dream but, for me, Lehrer's own attempt at a fourth culture fails. As Gerard Manley Hopkins might say, whatever "the achieve of, the mastery of" the artists, the science as laid out here seems "sheer plod", undermining the central conceit — for what artist would partake in such a paltry matter?

Lehrer's last lines posit that the individuality of our experiences is what science will never be able to serve, because "each of us literally inhabits a different brain, tuned to the tenor of our private desires". I was therefore surprised to find no mention of synaesthesia, whose recent scientific renaissance confronts the idiosyncratic perceptions described by artists such as Wassily Kandinsky, Alexander Scriabin and Vladimir Nabokov, and which potentially provides a fertile middle ground between art and science.

In *The Hidden Sense*, Cretien van Campen, a self-styled naturalist of synaesthesia, explores this ground, providing a gentle, insightful, often personal, account of coloured words, smells, tones, pains, even orgasms, that will fascinate scientists, artists, synaesthetes and others. Slowly, this deep, seemingly intractable, subjectivity yields to some phenomenology, simple questionnaires, elegant experiments, a dash of neuroimaging, and much thought and reflection.

Synaesthesia is not, says van Campen, "an audiovisual performance, a trend in art or music, a drug-induced hallucination, a kind of metaphor". Nor is it, he concludes, a rare, demarcated, distinctive anomaly like colour-blindness. Synaesthesia is instead the end of a continuum, manifesting as intermediate forms in most people. He classes synaesthesia within Kant's *sensus communis aestheticus*, emphasizing the pleasure, joy and "hypnotising beauty" of the phenomena experienced.

Chris McManus is professor of psychology and medical education in the Division of Psychology and Language Sciences at University College London, London WC1E 6BT, UK. His book *Right Hand, Left Hand* won the 2002 Aventis Prize for Science Books.

## Prions, pleasure and purple notes

### Proust was a Neuroscientist

by Jonah Lehrer

Houghton Mifflin: 2007. 256 pp. \$24

### The Hidden Sense: Synesthesia in Art and Science

by Cretien van Campen

MIT Press: 2007. 208 pp. \$29.95

### Chris McManus

"Oh no he wasn't!" might well be the response to Jonah Lehrer's claim that Proust was a neuroscientist. And neither were Walt Whitman, George Eliot, Auguste Escoffier, Paul Cézanne, Igor Stravinsky, Gertrude Stein or Virginia Woolf. For Lehrer, an editor on the science and culture glossy *Seed*, neuroscience has confirmed their artistic intuitions: "Proust was right about memory, Cézanne was uncannily accurate about the visual cortex, Stein anticipated Noam Chomsky, and Woolf pierced the mystery of consciousness."

Lehrer's conceit of the artist as a neuroscientist is not unique. Neurobiologist Semir Zeki feels Shakespeare and Wagner are "among the greatest of neurologists", probing the mind of man using language and music. Likewise, cognitive psychologist Patrick Cavanagh has written of painters whose use of shadows, colours, reflections and contours shows how the "visual brain uses a simpler, reduced physics", and hence that "artists act as research neuroscientists".

Call me old-fashioned, but neuroscience surely needs at least some indirect involvement with the greyish-pink stuff inside heads? However valid in their own way, painting, cooking, and writing novels, poems and music simply aren't neuroscience. Even cognitive science would be pushing it. Where are the methodology, the experimentation, the data and the hypothesis testing? "The impression is for the writer what experimentation is for the scientist," said Proust. But impressions are neither experiments nor science. The conceit remains

exactly that, if the term 'neuroscientist' is to retain any serious meaning.

The Proust chapter in Lehrer's elegant book includes some brief biographical background, a little on Henri Bergson's influence, and then the famous passage with the madeleine: "What did Proust learn from these prophetic crumbs of sugar, flour, and butter? He actually intuited a lot about the structure of our brain." These intuitions included "smell and taste are the only senses that connect directly to the hippocampus, the center of the brain's long-term memory, [whereas] all other senses are first processed by the thalamus, the source of language and the front door to consciousness". If indeed Proust intuited this anatomy, it was unfortunate because the taste pathway is wrong, and few regard the thalamus as the source of language — particularly Proust's compatriot, Paul Broca.

Neuroscientists are then painted as naïve and uncultured, assuming that memories are "just shelved away in the brain, like dusty old books in a library" and that the secret of memory lies in "senseless repetition". After a ramble through Cajal and Freud, and a rat experiment on remembrance and reconsolidation, comes a deep dive into neurobiologist Kausik Si's ideas on the prion-like cytoplasmic protein CPEB. This protein, Si believes, may participate in the formation of long-term memory, and in some way explains how Combray, the fictional village of Proust's childhood, could exist "silently below the surface, just behind the curtain of consciousness". Lehrer triumphantly concludes that "prions are, by definition, unpredictable and unstable, because memory obeys nothing but itself. This is what Proust knew: the past is never past".

The most interesting parts of *Proust was a Neuroscientist* are its manifestos on art and science in the prelude and coda. Lehrer accuses scientists of seeing the brain solely in terms of its physical details, as "nothing but a loom of



# Delving into the ancient brain

## On Deep History and the Brain

by Daniel Lord Smail

University of California Press: 2007.

286 pp. \$21.95, £12.95

## The Origins of Meaning: Language in the Light of Evolution

by James R. Hurford

Oxford University Press: 2007. 404 pp.

\$35

### Dean Falk

*On Deep History and the Brain* and *The Origins of Meaning* are both concerned to some degree with the evolution of human behaviour and the brain, but their similarity stops there. Historian Daniel Lord Smail takes on the full chronology of the human past and condenses it into a seamless narrative; James R. Hurford, a linguist, delves even further into the past to explore the evolutionary foundations of language.

Smail believes that history should bring prehistory into its fold, and focuses on biology, brain and behaviour in his endeavour. He first describes how the discovery and implications of deep time by geologists, biologists and naturalists in the mid-nineteenth century were the undoing of the sacred idea that humankind began relatively recently in the Garden of Eden. Historians then shifted from a sacred to a secular beginning — the rise of civilization in Mesopotamia. Thus, laments Smail, the Palaeolithic continued to receive short shrift and still needed to be 'historicized'. After all, humans who did not keep records still had a past. He has a point.

Smail examines the rupture that continues to separate prehistory from recorded history, together with the historiographical, epistemological and theoretical obstacles that have kept them apart. He explores the importance of biology in shaping cultural evolution, offering an interesting take on the nature/nurture dichotomy with his suggestion that Lamarckian mechanisms displaced Darwinian ones when human culture started to develop.

Smail's discussion of the brain is framed within a critique of evolutionary psychology, which considers that cognitive modules have been handed down, more-or-less unchanged from the deep evolutionary past — "Stone Age brains acting clumsily in modern environments." Cognition has been overemphasized, he believes, to the detriment of the drives, motivations, emotions, moods and other

neurophysiological states that are important for shaping human mental experiences.

The contention here is that certain cultural practices affecting brain-body chemistry through psychotropic mechanisms are amenable to a neurohistorical approach. Palaeolithic societies, for example, may have developed a range of mood-altering practices such as song, dance, ritual, and ingestion of mind-enhancing substances. Fast-forward to eighteenth-century Europe, when these were joined by caffeine, sentimental novels, pornography and a growing array of consumer goods. Today we have shopping, drugs and gossip.

Psychotropic practices have been used in the past by authoritarian institutions or governments to control people. Smail suggests that they arose slowly during the later Palaeolithic, accelerated in the Neolithic, and soared during the past few centuries; he attributes our enjoyment of them to the fact that we are social creatures who have always assessed our standing in groups through chemical cues. Perhaps so.

Although this is an enjoyable and creative book, it is not quite what I expected. There are no endocasts or sulcal patterns here, no Brodmann's area 10, or debates on brain size versus cortical reorganization (although Hobbits receive a brief mention). Notwithstanding these omissions, its flavour and neuro-historical

approach are original and provocative, and the suggestion rings true that the neurophysiological underpinnings of moods, motivations, and so on, were important during hominin cultural and neurological evolution.

Hurford's book, *The Origins of Meaning*, also addresses an interdisciplinary audience, extending to philosophers, psychologists, anthropologists, cognitive scientists and ethologists. The topic of language origins is an intellectual minefield, and those who tread on it need good balance — which Hurford has. Language, he neatly observes, allows humans to "go public with their thoughts".

A large part of the book reviews the experimental literature on the cognition of animals to model the protolinguistic substrates that probably characterized our earliest hominin ancestors. The late African grey parrot Alex, the famous vervet monkeys, and the chimpanzees Kanzi, Austin and Sherman are all here. So are a host of anonymous animals, such as bright piglets that back out of the 'wrong' arms of T-mazes, apparently to avoid giving incorrect responses. Do animals know what they know — that is, do they have self-awareness? Hurford shows that the seeds are there, and were probably present in our ancestors, providing fodder for natural selection.

Likewise, the capacity to realize something about objects (even that they are no longer present) is common to many animals and may have been a prerequisite for the emergence of the linguistic feature of 'displaced reference'. Hurford argues that the ability of many animals to infer animacy in objects (related to detecting biological motion) was necessary for our ancestors to evolve 'theory of mind'. Other comparative discussions include the evolution of episodic memory, which Hurford thinks emerged more recently than semantic memory, and the accurate perception of a number of objects without counting (usually around four), and the relationship of such abilities to the evolution of specific aspects of language.

Hurford locates an evolutionary starting point for language in early hominins in the simple two-way communications from which grammatical complexity and descriptive power eventually grew. He discusses physical and social environments of apes to locate possible precursors of group-wide reciprocal communications. He rejects male-female relationships and parent-offspring interactions because they are too asymmetrical, but views play behaviour among juveniles as more symmetrical and therefore a likely precursor. The literature on



Cranium of Tournai, the earliest known hominid. Will we ever know what went on inside it?

NATURE

mother–infant interactions, however, indicates that these are not asymmetrical and may have been a focus of intense natural selection.

The discussions in *The Origins of Meaning* of ritualization, emulation, imitation, learning and ‘machiavellian intelligence’ are insightful. So is the review of communications that include gaze-following, attention-seeking behaviour, pointing, begging and alarm calls in non-human primates and/or young humans, and their possible implications for the evolution of referential language. After considering niche construction, kin and sexual selection, and reciprocal altruism, Hurford concludes that communicative cooperation and trust may have been the most crucial factors in the emergence of language.

There are some titillating nuggets in this book, such as a discussion of how the *FOXP2* gene was mistakenly accepted as the ‘magic bullet’ responsible for language evolution. Even better is the extent to which academics from different countries use language competitively to show off — guess where Americans rank?

Has Hurford achieved his goal of describing the evolutionary foundations of language? Yes, elegantly and in accomplished detail that should be accessible to all those specialists the book targets. I look forward to finding out what he thinks happened next in the second volume. ■

Dean Falk is in the Department of Anthropology, Florida State University, Tallahassee, Florida 32306-7772, USA. She is the author of *Braindance: New Discoveries about Human Origins and Brain Evolution*.

## Space and time gift-wrapped

### Very Special Relativity: An illustrated Guide

by Sander Bais

Harvard University Press: 2007. 144 pp.  
\$20.95, £13.95

### Andrew Jaffe

Sander Bais's *Very Special Relativity* is a brief overview of Albert Einstein's 1905 theory of space and time, an esoteric topic that one would presume to be directed at a narrow audience. But this book would not be out of place on a coffee table, with its handsome design, thick, coloured papers and imaginative graphics that match the high standard of its content. We expect art books, not science texts, on our coffee tables, assuming that in science it is the ideas that matter, not the visual presentation.

Here, the elegant illustrations help Bais lead the reader from Einstein's postulates through the ideas of simultaneity, inertial frames, time dilation and relativistic energy and momentum, eschewing the usual admonitions against equations. The author's clever idea of pairing every page of text with a space–time diagram (a graphical tool actually used by relativists) to illustrate the concepts and mathematics suits the geometrical basis of its subject perfectly. It allows Bais to stop just short of using calculus — although the definitions of things such as ‘slope’ and ‘tangent’ are

taken for granted, and he does allow a  $dp/dt$  to slip in when discussing momentum and forces. He manages to navigate all the way to  $E=mc^2$  and to give us a taste of Einstein's even more miraculous work of the following decade, his general relativity theory of gravitation.

*Very Special Relativity* is aimed at an underserved market: keen high-school students will welcome it because it extends beyond qualitative discussion of ‘modern’ physics and popular-science books. With its brainteaser problems, it should also work as an undergraduate introductory textbook. Readers who haven't exercised their mathematical and geometric muscle since they were in school might find it tough going.

It is rare for science books to rate as objects in their own right, but *Very Special Relativity* is a lovely little object. You could easily imagine a web-based version of it, with a bit of animation to serve its pedagogical needs. Still, there is some quality about the hard covers and high resolution that even my 26-inch screen wouldn't be able to capture. No longer is there an excuse for physics textbooks to be expensive, boring, thick or stuffed with equations in order to qualify as good teaching material. ■

Andrew Jaffe is professor of astrophysics and cosmology in the Astrophysics Group, Blackett Laboratory, Imperial College London, London SW7 2AZ, UK.

### PAINTING

## Hidden depths spotted

### Emiliano Feresin

Nuclear physicists have revealed that the work of Sicilian artist Antonello da Messina (1430–79) may have been ahead of its time. He was among the first, they suggest, to subtly refine the shading in his paintings by spotting them with lacquer. When applied to particular areas of a painting, this glaze, made from red pigment mixed with oil, enhances the impression of depth.

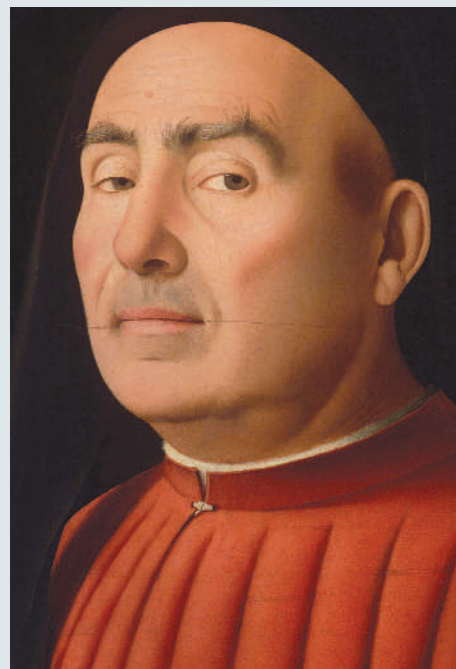
Pier Andrea Mandò and his colleagues at the Italian National Institute of Nuclear Physics' Labec laboratory in Florence took advantage of the availability of Antonello's *Portrait of a Man* (1476) during its restoration this summer. They had already developed and used non-destructive PIXE (for ‘particle-induced X-ray emission’) scans to reconstruct the spatial distribution and concentration of chemical elements on the surface of some types of artistic work, such as gold laces and embroideries. They wanted to try the technique for the first time on a painting.

The physicists identified traces of aluminium

and iron — a typical signature of lacquers — in tiny, uneven spots over the red cloak of the unknown subject. Although such glazes were already in use in the fifteenth century, they were normally applied with a brush over an entire surface. The scan reveals a weave-like imprint across the red cloak, suggesting to the restorers that here the shiny substance was dabbed on with a cloth.

Later, similar effects were created by great masters such as Leonardo da Vinci (1452–1519), who used his finger, and Rembrandt (1606–69), who used the end of his brush. But if Mandò and his colleagues are correct, it was Antonello — best known for introducing certain techniques of Flemish masters into Italy — who pioneered the effect. ■  
Emiliano Feresin is an intern in Nature's Munich office.

The restored *Portrait of a Man*, pictured here, can be viewed at the Civic Museum of Ancient Art in Palazzo Madama, Turin: [www.palazzomadamatorino.it](http://www.palazzomadamatorino.it).



CIVIC MUS. ANCIENT ART, TURIN (COURTESY OF TURIN MUSEUM FOUNDATION)



# Technologies of humility

Researchers and policy-makers need ways for accommodating the partiality of scientific knowledge and for acting under the inevitable uncertainty it holds.

**Sheila Jasanoff**

The great mystery of modernity is that we think of certainty as an attainable state. Uncertainty has become the threat to collective action, the disease that knowledge must cure. It is the condition that poses cruel dilemmas for decision-makers; that must be reduced at any cost; that is tamed with scenarios and assessments; and that feeds the frenzy for new knowledge, much of it scientific.

For a long time we accepted lack of certainty as humankind's natural lot. What has happened to reverse that presumption? Perhaps it is the spread of binary thinking that frames the future in terms of determinate choices between knowable options. Boolean algebra and digital logics are not only built into our computers, mobile phones and other information and communication technologies, they dominate the framing of social problems and the options for dealing with them.

Thus, statistics offers a choice between Type 1 and Type 2 errors. The first lead to false positives that promote too much risk avoidance, the second to false negatives that keep us from acting when we ought. Implicitly, error follows a binary trail. Philosophy casts moral dilemmas as trolley problems, in which possible solutions are represented as choices encountered at forks in the track. One option is to let the trolley run its course and let five people die; the other is to throw a fat man on to the track, diverting the trolley and killing only one person. Which is the moral choice? Decision theory adopts one way of thinking and reasoning as rational; all others are biased by definition and need to be explained away as aberrations of human cognition. Even the concept of the win-win solution assumes, in binary logic, that for each party to a game, winning and losing are the only options.

Life, as we know from experience, seldom unfolds in binaries. We rarely confront Hamlet's choice — to act or not to act. There are always added considerations. Which action is best, by what criteria, how soon, with what provisos, at what cost and with what allowance for error? Even the half-mad prince recognized that second-order consequences might complicate his first-order decision: to be or not to be.

Real problems in the real world are

infinitely complex, and for any given problem, science offers only part of the picture. Climate scientists can tell us with high certainty that human activities are raising Earth's mean surface temperature, that extreme weather events will occur, and that melting ice caps will cause abrupt



changes in the global climate. But it takes time and money to produce such certainty, and for all the doors that science even provisionally closes, others relevant to policy remain beyond closure by science alone. In the case of climate change, for example, science cannot tell us where and when disaster will strike, how to allocate resources between prevention and mitigation, which activities to target first in reducing greenhouse gases, or whom to hold responsible for protecting the poor. How should policy-makers deal with these layers of ignorance?

The short answer is with humility, about both the limits of scientific knowledge and about when to stop turning to science to solve problems. Policy-makers need to focus on when it is best to look beyond science for ethical solutions. And science advisers need to admit that other sorts of analyses must also inform political decisions. Capacity-building in the face of uncertainty has to be a multidisciplinary exercise, engaging history, moral philosophy, political theory and social studies of science, in addition to the sciences themselves.

Science fixes our attention on the knowable, leading to an over-dependence on fact-finding. Even when scientists recognize the limits of their own inquiries, as they often do, the policy world, implicitly encouraged by scientists, asks for more research. For most complex problems, the pursuit of perfect knowledge is asymptotic. Uncertainty, ignorance and indeterminacy are always present.

We need disciplined methods to accommodate the partiality of scientific knowledge and to act under irredeemable uncertainty. Let us call these the technologies of humility. These technologies compel us to reflect on the sources of ambiguity, indeterminacy and complexity. Humility instructs us to think harder about how to reframe problems so that their ethical dimensions are brought to light, which new facts to seek and when to resist asking science for clarification. Humility directs us to alleviate known causes of people's vulnerability to harm, to pay attention to the distribution of risks and benefits, and to reflect on the social factors that promote or discourage learning.

Policies based on humility might: redress inequality before finding out how the poor are hurt by climate change; value greenhouse gases differently depending on the nature of the activities that give rise to them; and uncover the sources of vulnerability in fishing communities before installing expensive tsunami detection systems.

This call for humility is a plea for policy-makers to cultivate, and for universities to teach, modes of knowing that are often pushed aside in expanding scientific understanding and technological capacity. It is a request for research on what people value and why they value it. It is a prescription to supplement science with the analysis of those aspects of the human condition that science cannot easily illuminate. It is a call for policy analysts and policy-makers to re-engage with the moral foundations for acting in the face of inevitable scientific uncertainty.

Sheila Jasanoff is Pforzheimer Professor of Science and Technology Studies, John F. Kennedy School of Government, Harvard University, 79 John F. Kennedy Street, Cambridge, Massachusetts 02138, USA.

For more essays and information see <http://nature.com/nature/focus/scipol/index.html>.

D. PARKINS

SCIENCE & POLITICS



## NEWS &amp; VIEWS

## SMELL

## The worm turns

Piali Sengupta

**The worm *Caenorhabditis elegans* has many advantages as an experimental organism. These have been exploited to investigate how, at a single-neuron level, neural circuits transform sensory signals into behaviour.**

Touching a hot stove elicits a highly predictable reflex action — you pull your hand away. But most complex behaviours are instead probabilistic in nature, being modulated by several variables, including the nature and intensity of incoming sensory signals and past experiences. How do neuronal circuits transform a deterministic stimulus into a probabilistic output? In this issue (page 63), Chalasani and colleagues<sup>1</sup> describe information flow in a neural circuit that drives a probabilistic motor output in response to olfactory cues in the nematode *Caenorhabditis elegans*.

This organism is particularly suited to analysing neural circuits (Fig. 1). It has relatively few neurons (around 300), which are readily identifiable and drive highly sophisticated behaviours<sup>2</sup>. But what makes this organism unique is that all the anatomical connections between its individual neurons are known<sup>3</sup>. In theory, this should allow researchers to follow a sensory stimulus through the circuit on a neuron-by-neuron basis, as it is transformed into a motor output. Chalasani *et al.* have put theory into practice, and measure neuronal activity in the pair of *C. elegans* olfactory neurons known as AWC and two of the neurons' downstream partners — the AIY and AIB interneurons — in response to odorant stimuli.

Previous work<sup>4–7</sup> had shown that, in response to the removal of attractive, food-related odours, AWC neurons increase the probability that the worm will assume a seeking behaviour, in which it turns for a sustained period of time, although the actual execution of a turn is stochastic, or random. Chalasani *et al.* used this system not only to investigate the generation of a probabilistic behaviour, but also to assess how a transient stimulus such as removal of food results in a prolonged change in behavioural output.

The authors first characterized the responses of the AWC neurons to odour stimuli, showing that exposure to an odour or the addition of food inactivates them, whereas removal of such stimuli activates them. The AIB interneurons respond like the AWC neurons, but the AIY interneurons behave in the opposite way — that is, they are activated in the presence of an odour and inhibited on odour removal (Fig. 2,



**Figure 1 | Model organism.** For several reasons, the worm *C. elegans* is researchers' favourite organism for analysing neural circuits.

overleaf). The responses of both AIBs and AIYs require the presence of AWC neurons.

As the AIY interneurons suppress turning behaviour in *C. elegans* and AIB interneurons enhance it<sup>4–6</sup>, these findings suggest that, on removal of an odour, the AWC neurons simultaneously activate AIBs and inhibit AIYs, thereby increasing the probability that the worm will turn. What happens on exposure to an odour is less clear, but it seems that the activation of the AIY interneurons contributes to the reduced probability of turning.

The devil is in the details. The authors note several crucial facts about the temporal dynamics of the responses that begin to explain how the incoming information is transformed by the circuit. They find that, although the response of the AWC neurons to odour removal is transient, the AIB response lasts longer (Fig. 2a). This observation may account for one mystery: how the 'working memory'<sup>8</sup> of a transient stimulus is maintained.

Chalasani *et al.*<sup>1</sup> also find that, once the primary response of AWC neurons decays after

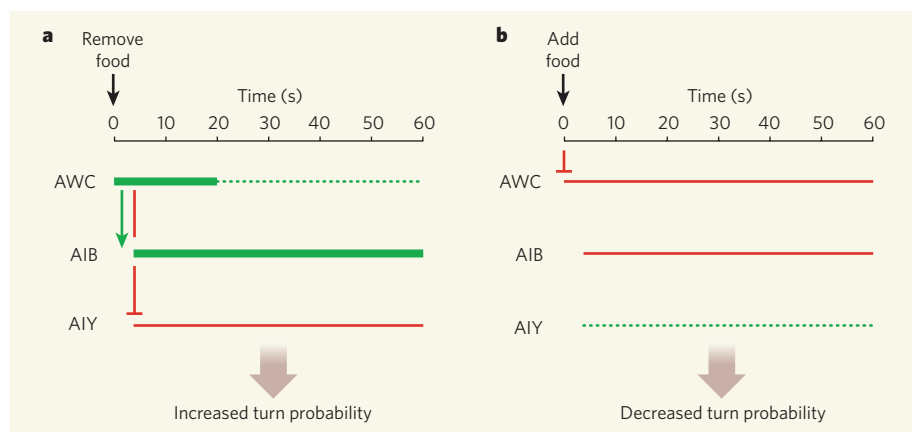
food removal, the neurons continue to respond in a stochastic manner for some time (Fig. 2a). Moreover, although AIY activity is stimulated by food odour, its response is transient, weak and stochastic (Fig. 2b). These observations indicate that the AIB neurons might 'gate', or act as a selective barrier to, turning, but that execution of an individual turn might also require contributions from the AWC and AIY neurons, as well as from other unidentified neurons. The findings also begin to explain both the stochastic nature of individual turns and their increased probability when food is removed and AIB neurons are activated. In this way, circuit activity can transform a deterministic stimulus into a probabilistic response.

Why not just respond in an all-or-none manner? The reason is that a circuit that generates probabilistic behaviour leaves itself open to further modulation. The circuit involving AWC, AIY and AIB neurons is only a sub-circuit within a much larger, interconnected neuronal network<sup>3</sup>. Inputs from this network could modulate the output of smaller sub-circuits, allowing the organism to integrate information from several environmental sources and accordingly alter the probability of a response. Thus, a probabilistic network might cope better with uncertainty and unpredictability — characteristics of the real-world environment.

An important next step will be to identify the molecules that mediate this information flow. Chalasani *et al.* have already made headway on this issue by showing that the AWC neurons release the neurotransmitter glutamate, which activates AIBs through a glutamate-gated cation channel and inhibits AIYs via a glutamate-gated chloride channel. But these analyses also indicate that other molecules are required to modulate olfactory responses.

An open question is the nature of the information flow downstream of the neurons studied here. Do the parallel activation and inhibition pathways remain segregated, or do they converge to regulate the output? What are the network junctions at which mechanisms such as experience- and feedback-dependent plasticity act to regulate circuit output? Sensory neurons carry out remarkably

MPG/EPA/CORBIS



**Figure 2 | Information flow in the olfactory circuit of *C. elegans*.** Chalasani *et al.*<sup>1</sup> show that in response to (a) removal and (b) addition of olfactory stimuli such as food, the AWC pair of neurons are activated or inhibited, respectively. This leads to the parallel, but opposite, regulation of two types of interneuron (AIB and AIY). The activity of this circuit determines the probability of seeking behaviour — turning — by the worm. Solid green lines indicate peak levels of activity and broken green lines show weaker, stochastic activity. Red lines indicate inhibition.

sophisticated computations<sup>9</sup>, implying that sensory circuits in *C. elegans* have a top-heavy design. Ultimately, simultaneous imaging of the activity of all the components of a circuit during information flow *in vivo*, and, better still, in freely behaving animals<sup>10,11</sup>, will be necessary if we are to understand the stability

and operation of neural networks<sup>12</sup>.

The work of Chalasani and colleagues<sup>1</sup> has identified a circuit motif that seems to be used in several contexts in different organisms. In vertebrate vision, information flow through parallel channels of opposite signs is crucial for contrast sensitivity<sup>13</sup>. The use

of a similar functional motif in *C. elegans* olfaction emphasizes the efficiency of this circuit in processing sensory information, and supports the idea that systems of very different complexities may nevertheless use shared strategies to perform similar tasks<sup>14</sup>. In other words, much within man might still be worm<sup>15</sup>.

Piali Sengupta is in the Department of Biology and the National Center for Behavioural Genomics, Brandeis University, Waltham, Massachusetts 02454, USA.  
e-mail: sengupta@brandeis.edu

- Chalasani, S. H. *et al.* *Nature* **450**, 63–70 (2007).
- de Bono, M. & Maricq, A. V. *Annu. Rev. Neurosci.* **28**, 451–501 (2005).
- White, J. G. *et al.* *Phil. Trans. R. Soc. Lond. B* **314**, 1–340 (1986).
- Tsalik, E. L. & Hobert, O. *J. Neurobiol.* **56**, 178–197 (2003).
- Gray, J. M., Hill, J. J. & Bargmann, C. I. *Proc. Natl Acad. Sci. USA* **102**, 3184–3191 (2005).
- Wakabayashi, T., Kitagawa, I. & Shingai, R. *Neurosci. Res.* **50**, 103–111 (2004).
- Hills, T., Brockie, P. J. & Maricq, A. V. *J. Neurosci.* **24**, 1217–1225 (2004).
- Dash, P. K. *et al.* *Learn. Mem.* **14**, 554–563 (2007).
- Clark, D. A., Biron, D., Sengupta, P. & Samuel, A. D. T. *J. Neurosci.* **26**, 7444–7451 (2006).
- Faumont, S. & Lockery, S. R. *J. Neurophysiol.* **95**, 1976–1981 (2006).
- Clark, D. A. *et al.* *J. Neurosci.* **27**, 6083–6090 (2007).
- Prinz, A. A., Bucher, D. & Marder, E. *Nature Neurosci.* **7**, 1345–1352 (2004).
- Schiller, P. H. *Trends Neurosci.* **15**, 86–92 (1992).
- Milo, R. *et al.* *Science* **303**, 1538–1542 (2004).
- Nietzsche, F. W. *Also sprach Zarathustra* (1885).

## ORGANIC CHEMISTRY

# Aromatics with a twist

Rainer Herges

**The properties of flat aromatic molecules are well known to chemists, but some non-planar aromatics remain a mystery. A molecule that can twist into a Möbius band on command might shed light on their features.**

The famous, one-sided Möbius strip can be constructed by cutting open a regular paper band, giving it a half-twist and then rejoining the ends. The same change in topology can now be performed at a molecular level — but without having to break open the loop. Reporting in *Angewandte Chemie*, Latos-Grażyński and colleagues<sup>1</sup> have prepared a molecule that switches from a regular to a Möbius band, simply by changing the polarity of the solvent. This isn't just a party trick — the change in shape transforms the electronic properties of the molecule.

The authors' work depends on the chemical concept of aromaticity. It seems likely that aromatic compounds acquired their evocative name because several of them have a pleasant smell — such as benzaldehyde (found in bitter almond oil), anethole (in aniseed) and cinnamaldehyde (in cinnamon). But this isn't always the case: benzene, the prototypical aromatic molecule, and its close relatives smell rather like gasoline. The real reason to group aromatic compounds together is that they have

similar chemical properties, which depend on their electronic structure.

The simplest aromatic compounds are known as annulenes. These are hydrocarbon rings that contain alternating single and double bonds (Fig. 1a). The double bonds are conjugated, which means that their electrons are not tied between two atoms and so are mobile, unlike the electrons in most other bonds. If a magnetic field is applied perpendicular to the ring plane of an annulene, it induces a 'ring current' in the mobile electrons. The current can be imagined as flowing in two circles above and below the ring, sandwiching the molecule like the filling in a bagel.

This simple model works surprisingly well in predicting the magnetic properties of annulenes. The direction of the ring current depends on the number of mobile electrons. For rings with  $4n+2$  electrons — where  $n$  is any integer greater than or equal to zero — the current circulates in a clockwise direction (assuming that the magnetic-field vector points towards the observer). But for annulenes with  $4n$

electrons (in this case,  $n$  is any integer greater than zero), the ring currents flow in the opposite direction.

In the 1930s, the physical chemist Erich Hückel predicted that annulenes containing  $4n+2$  mobile electrons should be stable (aromatic), and that those with  $4n$  electrons should be unstable (antiaromatic)<sup>2,3</sup>. His prediction spurred a number of synthetic chemists to prepare these molecules. By the end of the 1960s, most annulenes with up to 18 carbon atoms had been made, along with a few containing larger rings<sup>4</sup>. Each successful synthesis confirmed Hückel's predictions. The Hückel rule is now recognized as a fundamental law of chemistry. It allows the stability of conjugated ring systems to be predicted, and is something that every undergraduate chemist should know.

But in 1964, a spectacular theoretical prediction was made: the Hückel rules are reversed for annulenes with a 180° twist (so-called Möbius annulenes)<sup>5</sup>. High-level theoretical calculations confirmed that Möbius rings with  $4n$  electrons are indeed truly aromatic, whereas those with  $4n+2$  electrons are antiaromatic. Synthetic chemists once again rushed to prove the prediction experimentally, but all early attempts failed. This was because the twist in Möbius rings generates considerable molecular strain, the destabilizing effects of which could not be overcome by Möbius aromaticity. It wasn't until 2003 that the first synthesis of a Möbius annulene was achieved<sup>6</sup>, by inserting a well-designed structural unit into the ring to stabilize the twist. The properties of the resulting compound

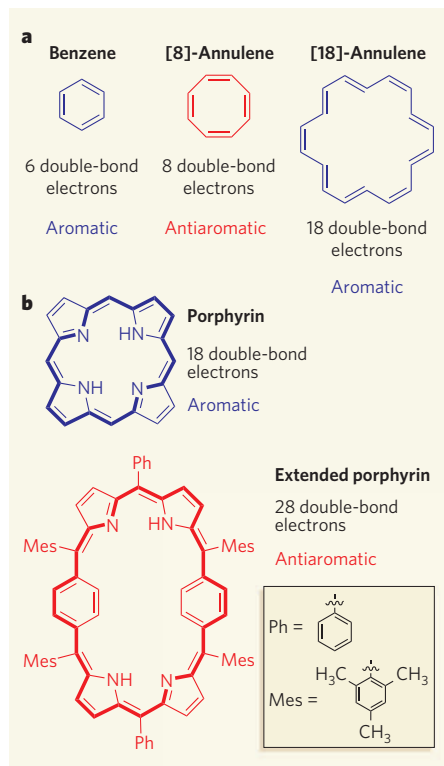


corroborated the controversial prediction.

And that was where the story ended — until now. Latos-Grażyński and colleagues<sup>1</sup> have synthesized a new kind of ring compound that exhibits Hückel or Möbius behaviour depending on the polarity of the surrounding solvent. Their molecule incorporates a 'porphyrin' structure (Fig. 1b). Porphyrins are aromatic molecules that are ubiquitous in nature — they form the reactive sites of haemoglobin, myoglobin and cytochrome enzymes, for example. Structures related to porphyrins are also found in chlorophyll and vitamin B<sub>12</sub>.

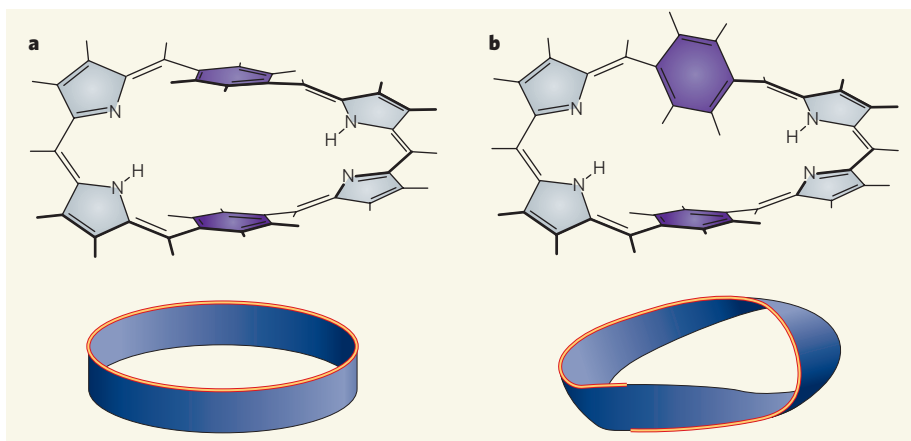
Latos-Grażyński and colleagues<sup>1</sup> have inserted two benzene rings into the porphyrin core (Fig. 1b), so increasing the number of mobile electrons from 18 (a  $4n+2$  system) to 28 (a  $4n$  system). Their nuclear magnetic resonance (NMR) experiments show that the compound is antiaromatic in nonpolar solvents, as expected from the Hückel rule. But in polar solvents, the antiaromaticity is lost. This change in electronic behaviour can be attributed to a molecular conformation change.

The difference in conformation seen in polar and nonpolar solvents is brought about



**Figure 1 | Annulenes and aromaticity.**

**a**, Annulenes are cyclic hydrocarbons that contain alternating single and double bonds, as shown in these examples. Annulenes with  $4n+2$  double-bond electrons (where  $n \geq 0$ ) are stable, and are described as aromatic compounds. Annulenes with  $4n$  double-bond electrons are unstable, and are described as antiaromatic compounds. **b**, The porphyrin structure is aromatic. Latos-Grażyński and colleagues<sup>1</sup> have made an extended porphyrin that is antiaromatic in nonpolar solvents. Only the bonds shown in bold affect the aromatic state of the molecule. The three-dimensional structure of the extended porphyrin is shown in Figure 2.



**Figure 2 | A molecular topological switch.** Latos-Grażyński and colleagues<sup>1</sup> have made a compound that is antiaromatic in nonpolar solvents, but not in polar solvents. **a**, In nonpolar solvents, the two benzene rings (purple) in the molecule are parallel, and the molecule is a two-sided, non-twisted band. **b**, In polar solvents, the upper benzene ring twists by  $90^\circ$ , so that the molecule becomes a one-sided, Möbius structure. This conformational change alters the aromaticity of the molecule.

by a small twist of one of the molecule's benzene rings, which acts like a topology switch (Fig. 2). If this benzene ring is perpendicular to the other one on the opposite side of the porphyrin, the molecule adopts a twisted, one-sided Möbius topology. But if the benzene rings are parallel to each other, the structure is untwisted and two-sided. The authors<sup>1</sup> did not observe aromaticity when the molecule was in the Möbius conformation, however, probably because the large amount of twist prevents efficient overlap of the electrons' orbitals.

This is the first example of a molecule with reversible topology that switches from an antiaromatic to a non-aromatic (or weakly aromatic) state. Naturally occurring porphyrin-containing compounds bind to metal cations to perform various functions, such as inducing redox reactions, transferring methyl ( $\text{CH}_3$ ) groups between molecules or storing oxygen. Artificial porphyrins that include a topology switch might extend the list of applications for these molecules.

When the first Möbius annulene was

prepared in 2003, it was predicted that additional molecules would be synthesized that would have stronger Möbius aromaticity<sup>7</sup>. We are still waiting for an example. But Latos-Grażyński and colleagues' work suggests that such a compound will probably be found among the extended porphyrins<sup>8</sup>.

Rainer Herges is at the Otto-Diels Institute for Organic Chemistry, University of Kiel, Otto-Hahn-Platz 4, D-24098 Kiel, Germany. e-mail: rherges@oc.uni-kiel.de

1. Stępień, M., Latos-Grażyński, L., Sprutta, N., Chwalisz, P. & Sztarbaum, L. *Angew. Chem. Int. Edn* **46**, 7869–7873 (2007).
2. Hückel, E. *Z. Phys.* **70**, 204–286 (1931).
3. Hückel, E. *Z. Phys.* **76**, 628–648 (1932).
4. Balaban, A. T., Banciu, M. & Ciorba, V. *Annulenes, Benzo-, Hetero-, Homo-Derivatives and their Valence Isomers* Vol. 1, 67–98 (CRC Press, Boca Raton, FL, 1987).
5. Heilbronner, E. *Tetrahedron Lett.* **5**, 1923–1928 (1964).
6. Ajami, D., Oeckler, O., Simon, A. & Herges, R. *Nature* **426**, 819–821 (2003).
7. Lemal, D. M. *Nature* **426**, 776–777 (2003).
8. Osuka, A. *et al.* 12th Int. Symp. Novel Aromatic Compounds, Awaji Island, Japan, 22–27 July 2007 *Pure Appl. Chem.* (in the press).

## VISION

# Dynamic platforms

Roger C. Hardie

**Scaffolding proteins are so named because they function as platforms for the assembly of molecular signalling complexes. But at least one such protein is more than a passive bystander and has its own signalling role.**

Living cells are dynamic machines. They contain thousands of different proteins that participate in a maze of signalling pathways, all of which must be accurately organized in space and time for efficient cellular functioning. To achieve this level of organization, one strategy that cells use involves scaffolding proteins, which assemble proteins of a particular

signalling pathway into multimolecular complexes, thereby promoting signalling specificity and speed while avoiding spurious crosstalk with other pathways<sup>1</sup>. Scaffolding proteins are often regarded as passive binding platforms. But in a paper published in *Cell*, Mishra *et al.*<sup>2</sup> show that INAD, a scaffolding protein essential for phototransduction in flies, undergoes

a light-induced conformational change, indicating that it has a direct and dynamic role in signalling.

Many scaffolding proteins contain one or more PDZ structural domains, which are modules of about 90 amino acids that bind to the carboxyl termini of specific protein partners<sup>2</sup>. PDZ-domain proteins have been implicated in almost all aspects of cell biology as structural binding platforms, but there are also reports<sup>3</sup> indicating that they have more direct signalling roles. Mishra *et al.* provide fascinating structural insight into such a role by solving the crystal structure of INAD — one of the best-characterized PDZ-domain scaffolding proteins<sup>4–8</sup> — in dynamically interchangeable states.

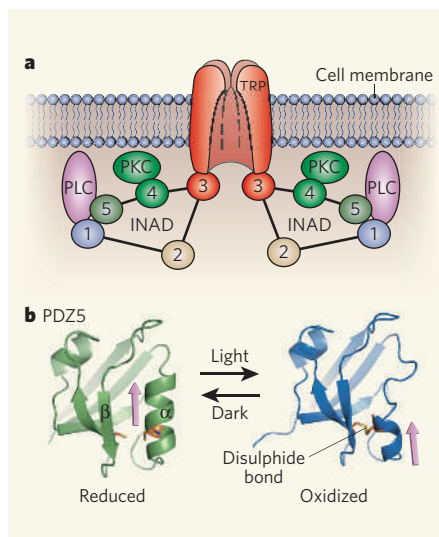
In the fruitfly *Drosophila*, INAD underpins the phototransduction cascade, a canonical, G-protein-coupled signalling pathway that can generate electrical responses within 20 milliseconds of the absorption of a single photon<sup>7,8</sup>. Crucial to the speed and amplification of this signalling cascade is the localization of its components within an array of about 30,000 tiny, finger-like membrane extensions, known as microvilli, that form the light-absorbing organelle in the photoreceptor cells.

Correct localization of the components of the phototransduction pathway depends on INAD, which has five interlinked PDZ domains that bind to different cascade components<sup>4–8</sup> (Fig. 1a). The core components of the resulting signalling complex include the enzyme required for excitation (PLC), the ion channel that mediates the electrical response to light (TRP) and a regulatory enzyme (PKC). PKC phosphorylates TRP and INAD, and possibly other proteins, and is required for the deactivation of the response to light<sup>6–8</sup>.

Using X-ray crystallography, Mishra *et al.*<sup>2</sup> have solved the structure of INAD's fifth PDZ domain (PDZ5) at 2 Å resolution. The ligand-binding site in PDZ domains is a groove between an  $\alpha$ -helix and a  $\beta$ -sheet, which is evolutionarily conserved in all 15 previously solved PDZ-domain protein structures<sup>9</sup>. Unexpectedly, the authors found that, owing to the formation of a disulphide bond, this ligand-binding groove becomes severely distorted in PDZ5 (Fig. 1b). In crystals grown under reducing conditions, however, they found that the disulphide 'bridge' was broken, and that PDZ5 had its expected topology.

The researchers then showed that, when flies are kept in darkness, PDZ5 is in its reduced form, but, after light stimulation, the disulphide bridge forms transiently, resulting in a conformational switch to the oxidized form. This 'redox switch' failed to operate in mutant flies lacking PKC, indicating that phosphorylation by this enzyme might trigger INAD's conformational change.

What is the functional significance of the redox switch in INAD's structure? Mishra *et al.* mutated one of the two cysteines in INAD that form the disulphide bridge (*Inad*<sup>C645S</sup>), thereby



**Figure 1 | Phototransduction in *Drosophila* and the INAD complex.** **a**, The five PDZ domains of INAD (1–5) assemble components of the phototransduction cascade, including PLC, the TRP channel and PKC, into a signalling complex at the cell membrane. **b**, Mishra *et al.*<sup>2</sup> report that, in response to light, the PDZ5 domain of INAD undergoes a conformational change. In the dark, PDZ5 is in its canonical, reduced form, in which a groove between an  $\alpha$ -helix and a  $\beta$ -sheet serves as a ligand-binding site. After stimulation with light, the PDZ5 domain undergoes a conformational change to an oxidized state, whereby the formation of a disulphide bond between two cysteine residues results in the unravelling of the  $\alpha$ -helix and the distortion of the ligand-binding groove. Following this conformational switch, the ligand (arrowed) — putatively part of the PLC enzyme — can no longer bind. (Adapted from ref. 2.)

effectively locking PDZ5 in the reduced state with its ligand-binding groove available. When tested with dim light, the photoreceptor cells of *Inad*<sup>C645S</sup> mutant flies showed normal responses. But, at higher light intensities, light responses deactivated slowly. The authors underlined the biological importance of the redox switch by showing that *Inad*<sup>C645S</sup> mutants have dramatic defects in visually mediated escape behaviour.

On the basis of these observations, Mishra *et al.* suggest a model according to which phosphorylation of INAD by PKC facilitates the transition of PDZ5 to the oxidized state. This results in the dissociation of the ligand — possibly PLC — from PDZ5, thereby inactivating and/or making PLC temporarily unavailable for renewed activation. But in *Inad*<sup>C645S</sup> mutant flies PLC would remain bound and could potentially be rapidly re-activated following the absorption of a second photon in the same microvilli, explaining the deactivation defect at high light levels.

This model is supported by earlier work<sup>10</sup> showing that PKC is required for rapid, calcium-dependent termination of PLC activity. As PLC is normally inactivated within milliseconds of  $\text{Ca}^{2+}$  influx into the cell, but can be

reactivated after a refractory period of about 100 milliseconds (ref. 7), this model implies that the reversible conformational switch in PDZ5 must be remarkably fast.

New advances always raise as many questions as they answer, and the study of Mishra *et al.* is no exception. Although earlier studies<sup>5</sup> concluded that PLC binds to PDZ5, Mishra *et al.* cite their own unpublished data questioning such an interaction. So, depending on whether PLC really is the ligand that binds to the PDZ5 domain of INAD, Mishra *et al.* suggest that their model may need to be refined.

The authors find that, although PDZ5 is stable in the oxidized state, when grown as crystals, the pre-stimulation, 'ground' state *in vivo* is the reduced form. How is the reduced state stabilized? Are there factors in the microvilli that create an unusually reducing environment, or do other (allosteric) factors favour the reduced state in the intact protein complex?

In addition, how does the activation of the phototransduction cascade facilitate the conformational switch? The requirement for PKC suggests phosphorylation of INAD may trigger the switch. But INAD is not the only PKC target. Light stimulation also rapidly initiates dramatic changes in the local (microvillar) environment. These include  $\text{Ca}^{2+}$  influx through TRP channels, which leads to exceptionally high  $\text{Ca}^{2+}$  concentrations, and huge changes in the levels of the PLC substrate  $\text{PtdInsP}_2$  and its metabolites DAG and  $\text{InsP}_3$ . Whether light exposure also induces changes in the redox state within the microvilli is not known.

Finally, how common is this new mechanism? The specific cysteine residues involved in the formation of the disulphide bridge seem to be unique to the PDZ5 domain of INAD in flies that belong to the Brachycera suborder of the order Diptera. This suggests that the redox switch may be a specific adaptation for the rapidly signalling visual systems of these organisms. However, Mishra and colleagues' graphic demonstration<sup>2</sup> of a dynamic conformational switch and its functional significance will surely motivate the search for similar dynamic behaviour among other members of the large and ancient family of PDZ-domain proteins. ■

Roger C. Hardie is in the Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge CB2 3DY, UK.  
e-mail: rch14@hermes.cam.ac.uk

- Sheng, M. & Sala, C. *Annu. Rev. Neurosci.* **24**, 1–29 (2001).
- Mishra, P. *et al. Cell* **131**, 80–92 (2007).
- Kim, E. & Sheng, M. *Nature Rev. Neurosci.* **5**, 771–781 (2004).
- Shieh, B.-H. & Zhu, M.-Y. *Neuron* **16**, 991–998 (1996).
- Tsunoda, S. *et al. Nature* **388**, 243–249 (1997).
- Huber, A. *Eur. J. Neurosci.* **14**, 769–776 (2001).
- Hardie, R. C. & Raghu, P. *Nature* **413**, 186–193 (2001).
- Wang, T. & Montell, C. *Pflugers Arch.* **454**, 821–847 (2007).
- Doyle, D. A. *et al. Cell* **85**, 1067–1076 (1996).
- Gu, Y., Oberwinkler, J., Postma, M. & Hardie, R. C. *Curr. Biol.* **15**, 1228–1234 (2005).



## MATHEMATICAL PHYSICS

## First encounters

Michael F. Shlesinger

**The idea of 'random walks' pops up in areas from biochemical reaction pathways to animals' foraging strategies. A central question — how likely is it that a walker is somewhere for the first time? — now has a simpler answer.**

A walker sets out on a random walk, moving between many different sites in no particular order. Formally, knowing the probability of arriving at each of the sites for every possible number of steps provides all the information required to characterize the walk. But it doesn't necessarily give easy answers to some relevant questions: for instance, what is the probability that the walker will reach a certain site for the first time after a given number of steps? This 'first-passage time' problem is notoriously intractable. For that reason, we often settle for the somewhat simpler problem of calculating the average (mean) first-passage time to an arbitrary site.

On page 77 of this issue, Condamin *et al.*<sup>1</sup> revisit the question of the mean first-passage time, and derive some new results of surprising simplicity. The authors consider a finite space made up of a lattice of nodes. This lattice can be arbitrarily regular, disordered or hierarchical, or it can assume some configuration, such as that of a scale-free network. The Internet,

and many models of human social interactions, are good examples of the scale-free case. They consist of nodes with many different degrees of connectedness, and the likelihood that a node has a certain number of links declines as a power law with that number: many people (or network nodes) have a small number of connections, but very few have a very large number. First-passage time distributions thus have relevance, for example, for the spread of sexually transmitted diseases in a human social network, or the spread of computer worms through the Internet.

The authors find three cases for the mean first-passage time between an arbitrary source and a designated target at a distance  $r$ . To a good approximation, this time always grows linearly with the number of nodes,  $N$ , but also depends either on  $r$  raised to a certain power or, in a special case, the logarithm of  $r$  (we'll return to what these three cases signify in a moment). The linear dependence on  $N$  seems to contradict results calculated by

Elliott Montroll<sup>2</sup> almost 40 years ago. He was studying the random walk of an exciton (an electron-hole excitation carrying energy) towards a reaction site in a photosynthetic plant cell, and found that, assuming jumps are to nearest-neighbour sites, the mean first-passage time differs according to whether the sites are arrayed in a line, a square lattice or a simple cubic lattice. (The exact formulae are  $N^2/6$ ,  $(1/\pi)N/\ln N$  and  $1.5N$ , respectively.)

The reason for the apparent discrepancy is that Montroll averaged over all possible starting points for the walk, thus removing any dependence on  $r$  from his problem. Similarly, if every one of the nodes could be visited on any jump with equal probability, the node structure and connections would be irrelevant and the mean first-passage time would be exactly equal to  $N$ . It is in fact the node connections that generate the dependence on  $r$ . Because Condamin *et al.*<sup>1</sup> explicitly take account of the role of the source-target separation in their formulation of the problem, their results are more fundamental and, when averaged over the possible  $r$  values, Montroll's expressions pop out again.

Condamin *et al.* parametrize the mean first-passage time in terms of two characteristic 'dimensions'. These are not dimensions in the usual sense of one, two or three, but are derived from scaling relationships between the variables that are inherent in the problem. First up is the 'fractal dimension',  $d_f$ , which characterizes the density of the nodes in the lattice. In a lattice of two dimensions, the meaning of this quantity can be envisaged by drawing a circle enclosing part of the plane of the lattice, and counting the number of nodes it encloses. Then you increase the radius of the circle, and calculate again, repeating this process several times. The number of enclosed nodes will scale as some power of the circle's radius; that power is  $d_f$ . For a lattice with a constant average density of nodes in the plane,  $d_f$  will be 2 — the number of enclosed points scales with area. But for fractal shapes in the plane, such as a Sierpinski gasket (Fig. 1a on page 78) or a so-called percolation cluster, which have a hierarchy of self-similar node groupings that only sparsely fill the plane,  $d_f$  will be less than 2.

The second dimension in the authors' formulae<sup>1</sup> is the 'walk dimension',  $d_w$ , which characterizes the ease with which the walker can move through the lattice. Mathematically, it enters the expression for the mean-square displacement from the source as a function of the number of moves,  $n$ , which is written to scales as  $n^{2/d_w}$ . On our regular two-dimensional lattice,  $d_w$  would be equal to 2 (meaning that the mean-square displacement grows linearly with the number of moves made, the 'diffusive' case). But on a lattice with many bottlenecks and dead-ends, the walk can be severely impeded, producing a  $d_w$  of greater than 2, and a process slower than diffusion.

It is the difference between the fractal



**Have I been here before?** Condamin and colleagues' mathematical results<sup>1</sup> deal with the likelihood that a walker who has embarked on a random walk is visiting a place for the first time.

ACE STOCK LTD/ALAMY

dimension and the walk dimension that determines which of Condamin and colleagues' three cases applies. If the walk dimension is greater than the fractal dimension ( $d_w > d_f$ ) — if the walker is significantly impeded or the lattice is sparse — a compact, local exploration arises. In this case, the authors find a strong, positive dependence of the mean first-passage time on the initial source–target separation, of the form  $r^{d_w-d_f}$ . They calculate  $d_w$  and  $d_f$  exactly for the case of a Sierpinski fractal lattice, and find good agreement between the predicted and calculated dependencies of the mean first-passage time on both  $N$  and  $r$ .

The opposite case, in which the fractal dimension is greater than the walk dimension ( $d_f > d_w$ ), characterizes a non-compact exploration, in which a random walker leaves a region with many sites still unvisited. The Lévy flight is one such case. This is a random walk that

performs jumps of many sizes, making its starting point irrelevant. It has been proposed as an optimal search strategy for, say, animals foraging for sparse resources. Here, the authors' formula for the mean first-passage time also shows a dependence on the source–target separation of the form  $r^{d_w-d_f}$ . Because the exponent is by definition negative, however, this dependence disappears at large values of  $r$ , leaving the mean first-passage time dependent on the number of nodes alone. Finally, in the authors' third case,  $d_w = d_f$ , they find a weak, logarithmic dependence of the mean first-passage time on the separation.

The authors' methods and calculations<sup>1</sup> cut to the essence of the problem of the mean first-passage time with a simple, general solution in terms of just the number of nodes in the lattice on which the walk takes place, and the separation of target and source. But their work closes a

chapter, not a whole book: many other types of problem involving first-passage times fall outside the terms on which this model was based. This is seen, for example, in the search strategies of some predatory animals, which mix elements of the random-walk model with concentration on seasonal 'hot spots' to find prey. Similarly, in biological cells, peptide binding to transmembrane receptors depends on hydrophobic attraction superimposed on thermodynamic, random brownian motion. Such situations will continue to provide an unlimited number of questions for the mathematician. ■

Michael F. Shlesinger is in the Office of Naval Research/Code 30, 875 North Randolph Street, Arlington, Virginia 22203, USA.  
e-mail: mike.shlesinger@navy.mil

1. Condamin, S., Bénichou, O., Tejedor, V., Voituriez, R. & Klafter, J. *Nature* **450**, 77–80 (2007).
2. Montroll, E. W. *J. Math. Phys.* **10**, 753–765 (1969).

## PLANT PATHOLOGY

# Deadly special deliveries

Nicholas J. Talbot

**When attacking a plant, pathogens must deliver proteins into their victim's cells. The causal agent of potato late blight uses a system that is remarkably similar to that used by the malaria parasite in red blood cells.**

To infect plants and cause disease, many microorganisms evade or subdue plant defences so that they can proliferate unhindered within the host's tissues<sup>1</sup>. For this purpose, pathogenic bacteria have systems to deliver 'effector' proteins directly into plant cells, where they interact with plant proteins and suppress defence mechanisms<sup>2</sup>.

Whisson *et al.* (page 115 of this issue<sup>3</sup>) describe how the agent of potato late blight, *Phytophthora infestans*, uses a special host-cell-targeting signal<sup>4</sup> to attack its plant host. This provides the first clue as to how pathogen proteins are delivered into plant cells by a eukaryote — the vast group of organisms that differ from bacteria in having cells with membrane-bound nuclei. Beyond that, it turns out that *P. infestans* uses a signal that closely resembles that used by the malaria parasite, *Plasmodium falciparum*, to deliver parasite proteins into red blood cells<sup>4–6</sup>.

There has been rapid progress in understanding how bacteria attack both animals and plants using an array of effector proteins<sup>2,7</sup>. Bacteria deliver effectors mainly by use of a mechanism, called the type III secretion system, that allows proteins to be sent directly into the cytoplasm of host cells<sup>7</sup>. By contrast, we know little about effector proteins in the plant pathogenic fungi and oomycetes such as *P. infestans*. Oomycetes physically resemble fungi, but are in fact closely related to brown algae, and are responsible for some

of the most devastating plant diseases.

*Phytophthora infestans* attacks potato plants, and was the cause of the Irish potato famine of 1845–49 that resulted in the death of up to a million people and the emigration of as many again<sup>8</sup>. Whisson *et al.*<sup>3</sup> investigated how a particular effector protein, Avr3a, is delivered into potato plants. Avr3a is one of many secreted proteins in *P. infestans* that contain a group of amino acids with the sequence (in single-letter code) RXLR-EER — with X being any amino acid, and EER occurring within 25 amino acids of the RXLR motif<sup>9</sup>. This motif (both with and without EER) has been identified in Avr proteins of other oomycete pathogens, such as *Hyaloperonospora parasitica*<sup>9</sup>. It also resembles the RXLXE/Q motif found in the malaria pathogen *P. falciparum*, which belongs to a group of animal parasites known as the Apicomplexa.

In *P. falciparum*, the RXLXE/Q motif is necessary for the translocation of proteins into red blood cells<sup>4–6</sup>. After entering the red blood cell, the parasite occupies a compartment called the parasitophorous vacuole and secretes substances into the vacuole, from where RXLXE/Q-containing proteins are then selectively delivered into the cytoplasm of the red blood cell<sup>4–6</sup> (Fig. 1a, overleaf). It was known that the RXLR-EER motif from oomycetes can operate as the host-cell-targeting signal in *P. falciparum*, which suggested that there might be a common delivery mechanism for

oomycete and apicomplexan effectors<sup>10</sup>. Until now, however, there has been no direct experimental evidence that the RXLR-EER sequence is necessary for protein delivery in plant cells.

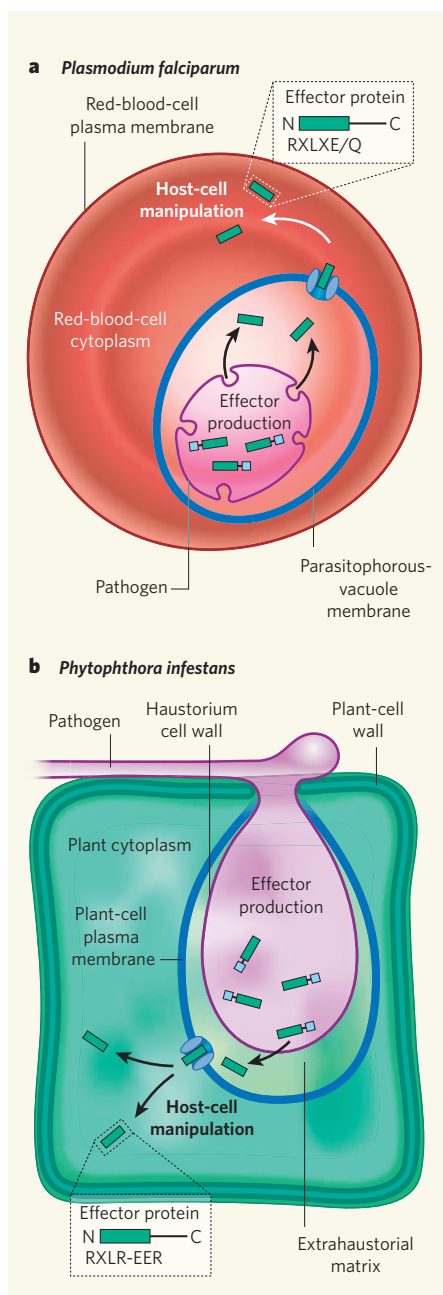
The Avr3a effector probably suppresses plant defences. Plants have evolved resistance proteins to counter this sort of intracellular microbial attack, however, and inside a cell these proteins interact (directly or indirectly) with Avr proteins to bring about cell death and thereby prevent further infection<sup>1</sup>. Expression of Avr3a inside plant cells showed that the RXLR-EER motif in itself is not necessary for Avr3a to induce a host response<sup>3,9</sup>. But the motif is necessary for the activity of Avr3a when it is secreted by *P. infestans*, implying that it is essential for host-cell targeting<sup>9,10</sup>.

To test this idea, Whisson and colleagues<sup>3</sup> replaced RXLR-EER with alanine residues or with the sequence KMIK-DDK, both of which maintained the predicted structure of Avr3a but resulted in loss of the potential host-cell-targeting signal. In these cases, the mutations prevented *P. infestans* from provoking a host response, suggesting that delivery of Avr3a to its site of action was not happening.

To investigate matters further, the authors fused the gene encoding Avr3a with a gene encoding  $\beta$ -glucuronidase, an enzyme known to work only inside plant cells<sup>11</sup>. When this reporter gene was expressed in *P. infestans*, the delivery of Avr3a to plant cells was observed directly using an enzyme assay that leads to a coloured product, which could be seen clearly in plant cells. This experiment provides convincing evidence that RXLR-EER acts as a host-cell-targeting sequence.

Analysis of the *P. infestans* genome shows that it may encode as many as 425 proteins with the RXLR-EER motif, 169 of which are very strong candidates<sup>3</sup>. These proteins are likely to act as effectors both for modulating the host response and for implementing the structural alterations in host cells necessary for invasion





**Figure 1 | Effector-protein delivery systems in *Plasmodium falciparum* and *Phytophthora infestans*.** **a**, After entry into a red blood cell, *Plasmodium falciparum* delivers host-cell-targeting effector proteins with the RXLXE/Q motif into the parasitophorous vacuole. These are then taken up into the cytoplasm of the red blood cell. **b**, *Phytophthora infestans* forms a specialized feeding structure known as a haustorium, which is the site of effector-protein secretion into the plant cell through the RXLR-EER host-cell-targeting system studied by Whisson and colleagues<sup>3</sup>. Proteins are delivered into the extrahaustorial matrix and then cross the plant-cell plasma membrane into the cytoplasm of the plant cell. In both cases a putative ATP-dependent translocator protein is shown. The process could, however, also operate through endocytosis. The blue box on the protein at the site of effector production depicts a signal sequence that is cleaved during secretion from the pathogen.

by the feeding structures, known as haustoria, that *P. infestans* produces inside plant cells (Fig. 1b). Identifying these effectors and their targets will provide insight into the processes necessary for the initial spread of the pathogen in plant tissue.

Another task will be to work out the RXLR-EER-mediated mechanism of host-cell targeting; it is likewise not yet clear how the RXLXE/Q signal operates to allow *P. falciparum* proteins to cross the membrane of the parasitophorous vacuole. Exploiting the power of plant genetics — especially in the model plant *Arabidopsis thaliana*, which is the host of *H. parasitica* — may allow rapid progress to be made in tackling this problem.

Finally, the presence of such a similar effector-delivery system in such different organisms as an apicomplexan and an oomycete prompts evolutionary questions. What is the likely origin of this system, and how widespread is

it? Does it exist in relatives of the oomycetes, which include the human parasite *Blastocystis*, or in even more disparate groups of eukaryote pathogens?

Nicholas J. Talbot is in the School of Biosciences, University of Exeter, Geoffrey Pope Building, Stocker Road, Exeter EX4 4QD, UK.  
e-mail: n.j.talbot@exeter.ac.uk

1. Jones, J. D. G. & Dangl, J. L. *Nature* **444**, 323–329 (2006).
2. Grant, S. R., Fisher, E. J., Chang, J. H., Mole, B. M. & Dangl, J. L. *Annu. Rev. Microbiol.* **60**, 425–449 (2006).
3. Whisson, S. C. et al. *Nature* **450**, 115–118 (2007).
4. Przyborski, J. & Lanzer, M. *Science* **306**, 1897–1898 (2004).
5. Hiller, N. L. et al. *Science* **306**, 1934–1937 (2004).
6. Marti, M., Good, R. T., Rug, M., Knuepfer, E. & Cowman, A. F. *Science* **306**, 1930–1933 (2004).
7. Galán, J. E. & Wolf-Watz, H. *Nature* **444**, 567–573 (2006).
8. Boyle, P. P. & Ó Gráda, C. *Demography* **23**, 543–562 (1986).
9. Birch, P. R. J., Rehmany, A. P., Pritchard, L., Kamoun, S. & Beynon, J. L. *Trends Microbiol.* **14**, 8–11 (2006).
10. Bhattacharjee, S. et al. *PLoS Pathogens* **2**, e50 (2006).
11. Denecke, J., Botterman, J. & Deblaere, R. *Plant Cell* **2**, 51–59 (1990).

## HEARING

# A fantasia on Kölliker's organ

Ian D. Forsythe

**In the silence that precedes the onset of hearing in the developing auditory system, it seems that the cells of a transient structure known as Kölliker's organ are capable of generating their own 'virtual' music.**

From a physiological perspective, a developing organ requires a programme that allows it to grow and adapt to internal and environmental constraints. In sensory systems such as those involved in sight and hearing, the adaptable growth of afferent (incoming) nerve fibres is involved in connecting the peripheral sensory organ to the neurons of the central nervous system. Information from the sense organ passes along the afferent nerve fibres in the form of electrical action potentials.

Sperry's chemoaffinity hypothesis<sup>1</sup> proposes that nerve path-finding requires the presence of guidance molecules on the target cells and the growing nerves, and this implies that an internal, or genetic, programme is the main determinant of development. But it is becoming increasingly apparent that sensory information in the form of electrical activity provides complementary, experience-related influences<sup>2</sup>. Thus, the growth of the afferent nerves from the developing sense organs to their targets in the brain begins with genetically encoded guidance, but over time is increasingly affected by experience-dependent processes. In a paper published in this issue (page 50), looking at the auditory system, Tritsch et al.<sup>3</sup> extend the debate on the importance of experience-dependent development to times before the ear can hear. And they identify a new, sound-independent means of generating neuronal activity.

The ear is a complex organ consisting of outer, middle and inner parts (Fig. 1a, overleaf). After passing through the outer and middle ear, sound reaches the base of the cochlea, a conical-shaped structure in the inner ear. The sound energy induces a travelling wave that propagates along the length of the cochlea, causing the overlying organ of Corti to vibrate at different longitudinal positions depending on the frequencies contained in the sound (low-frequency sounds resonate near the apex of the cochlea and high-frequency sounds resonate at the base). This amplified resonance flexes microscopic hairs (stereocilia) on adjacent inner hair cells (IHCs) and is converted into electrical potentials. Thus, the position along the cochlea predicts the frequency of sound that will stimulate each IHC. This 'place map' is referred to as tonotopy.

The neurotransmitter glutamate is then released from the IHCs through a calcium-dependent secretory pathway, and activates receptors on the afferent nerve fibres that innervate IHCs. This activation produces action potentials (with peak sound-evoked firing rates of more than 300 hertz) that are conducted along the eighth cranial nerve to the brain<sup>4</sup>. So one might imagine a symphony (or organ recital) encoded by the cochlea, with harmonies of waxing and waning action potentials in a parallel array of 40,000 afferent nerve fibres — the high notes augmenting activity in

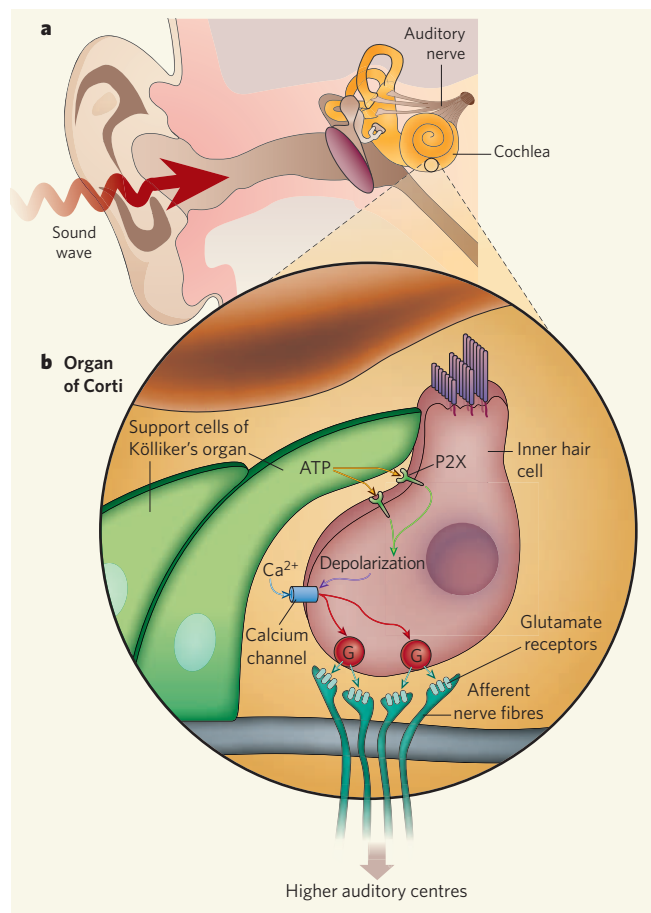
the basal nerves and the bass increasing the apical-nerve activity.

But in the absence of sound, the afferent nerve fibres are not silent, because IHCs continue to release low levels of glutamate, triggering spontaneous action potentials. In mature animals, this spontaneous activity can range from less than 1 Hz to more than 150 Hz in different afferent fibres, allowing upwards or downwards modulation of the firing rates. With so much electrical activity, it is not surprising that the auditory centres have the highest cellular metabolic rates in the brain.

It had been considered that, before the onset of hearing in an animal, which is thought to coincide with the opening of the auditory canal (in rats and mice, at around 11–12 days after birth), spontaneous activity substituted for sound-evoked activity. If spontaneous activity occurred randomly, it would be less likely to trigger experience-dependent adaptive processes. But in the immature animal it is clustered<sup>5</sup>, and this is consistent with the periodic activation of groups of afferent fibres by some unknown mechanism in the absence of sound input.

Tritsch *et al.*<sup>3</sup> have found that this early spontaneous activity originates from non-neuronal support cells in a transient structure of unknown function called Kölliker's organ located adjacent to the IHCs and running the length of the organ of Corti (see Fig. 1a on page 51). These cells release the chemical messenger ATP, which spreads locally, binding to P2X and P2Y receptors on adjacent cells and cascading onto the IHCs and their afferent fibres (Fig. 1b). The authors' finding that ATP release does not occur through the conventional calcium-mediated secretory pathway, but instead seems to be mediated by 'gap junctions', is important. Gap junctions usually allow small molecules, such as potassium, calcium and ATP, to pass between cells, so they might also allow ATP out of the support cells. Mutations in proteins that form these gap junctions (such as connexin 26 and connexin 30) are linked with the most common types of congenital deafness<sup>6</sup>. It will therefore be interesting to see whether these mutations affect the function of Kölliker's organ.

Tritsch and colleagues carry out recordings from single cells and non-invasive optical recordings from Kölliker's organ to show both spontaneous activity and raised intracellular calcium levels in clusters of cells along the organ of Corti. They then show how ATP release from the support cells excites local IHCs to release glutamate, which then triggers bursts of action potentials in auditory nerve fibres



**Figure 1 | Supporting the auditory pathway.** **a**, After passing through the outer and middle ear, sound waves reach the cochlea in the inner ear. **b**, In the mature ear, sound vibration causes depolarization of the membrane of the inner hair cells. This opens voltage-gated calcium channels, allowing  $\text{Ca}^{2+}$  to flood into the cell to trigger release of the neurotransmitter glutamate. Tritsch *et al.*<sup>3</sup> find that, in the immature ear, Kölliker's organ releases ATP, which binds to P2X receptors on inner hair cells to cause depolarization and calcium influx, so mimicking the effect of sound input. In both cases, glutamate activates receptors on the afferent fibres, triggering electrical action potentials that propagate along the nerve fibres to the brain.

(Fig. 1b). The authors' achievement in recording directly from the very fine nerve fibres in contact with the base of the IHCs, and showing how the fibres are excited, is an impressive feat of cellular physiology.

So what do these results mean for our understanding of hearing? A prerequisite for experience-dependent adaptation is that the spontaneous activity should be elicited in a coherent or simultaneous manner, thereby defining a related population of nerve fibres. The observed synchronized activity in IHCs across a distance of around 60 micrometres, or 6–10 IHCs (and desynchronization between more distant IHCs), supports the idea that this activity may have a signalling function in defining the association between adjacent regions of the organ of Corti (tonotopy). Inevitably, this activity would cascade onto each subsequent higher level of auditory processing, moulding the development of the central auditory pathways and refining connectivity between the nerve-cell junctions, or synapses<sup>5</sup>.

This process is important because in

sensory regions of the brain, the afferent nerve fibres and their contacts with their target neurons maintain a topographic relationship with the peripheral sense organ through chemoaffinity mechanisms, which involve guidance molecules, and experience-dependent refinement<sup>2,5,7</sup>. For example, a topographic representation of two-dimensional visual space on the retina (referred to as retinotopy) is maintained at higher levels of the visual pathway and serves to guide visual processing. In the cochlea, sound frequency is mapped along the single row of IHCs running the length of the organ of Corti, and this tonotopic map is preserved as the afferent fibres in contact with the IHCs grow into the brain. An important goal will be to explore the ways in which activity-dependent processes interact with the chemoaffinity and neurotrophic mechanisms in the developing auditory pathway.

The work of Tritsch *et al.* highlights the need to re-examine activity-dependent modulation in developing central auditory pathways at much earlier times than were previously thought relevant. These fundamental studies provide insights not only into the mechanisms underlying auditory development, but also into those involved in deafness, tinnitus and implant therapies. Indeed, tonotopy is the principle behind cochlear implants that directly stimulate the eighth-nerve fibres in a deaf person's cochlea, and it permits implant technology to be

extended to the cochlear nucleus and midbrain in patients with degeneration of the auditory nerve itself.

Ian D. Forsythe is at the MRC Toxicology Unit, Hodgkin Building, University of Leicester, Lancaster Road, Leicester LE1 9HN, UK. e-mail: idf@le.ac.uk

1. Sperry, R. W. *Proc. Natl. Acad. Sci. USA* **50**, 703–710 (1963).
2. Cline, H. *Trends Neurosci.* **26**, 655–661 (2003).
3. Tritsch, N. X., Yi, E., Gale, J. E., Glowatzki, E. & Bergles, D. E. *Nature* **450**, 50–55 (2007).
4. Geisler, C. D. *From Sound to Synapse* (Oxford Univ. Press, 1998).
5. Rubel, E. W. & Fritschsch, B. *Annu. Rev. Neurosci.* **25**, 51–101 (2002).
6. Petit, C. *Trends Mol. Med.* **12**, 57–64 (2006).
7. Cramer, K. S. *Hearing Res.* **206**, 42–51 (2005).

#### Correction

In the News & Views article "Linguistics: An invisible hand" by W. Tecumseh Fitch (*Nature* **449**, 665–667; 2007), artist's errors occurred in Figure 1, the glossogenetic tree of Indo-European language redrawn from a historical source. "Islamic" should read "Slavic". "Ukranian" should read "Ukrainian". And the branch leading to the Greek group of languages should likewise be labelled "Greek".



# Patterns of relative species abundance in rainforests and coral reefs

Igor Volkov<sup>1,2</sup>, Jayanth R. Banavar<sup>1</sup>, Stephen P. Hubbell<sup>3,4</sup> & Amos Maritan<sup>5</sup>

**A formidable many-body problem in ecology is to understand the complex of factors controlling patterns of relative species abundance (RSA) in communities of interacting species. Unlike many problems in physics, the nature of the interactions in ecological communities is not completely known. Although most contemporary theories in ecology start with the basic premise that species interact, here we show that a theory in which all interspecific interactions are turned off leads to analytical results that are in agreement with RSA data from tropical forests and coral reefs. The assumption of non-interacting species leads to a sampling theory for the RSA that yields a simple approximation at large scales to the exact theory. Our results show that one can make significant theoretical progress in ecology by assuming that the effective interactions among species are weak in the stationary states in species-rich communities such as tropical forests and coral reefs.**

A variety of patterns have been observed in the RSA distributions, which are measures of the number of species having a given number of individuals, of ecological communities. In particular, tropical forests<sup>1–5</sup> and coral reefs<sup>6,7</sup> exhibit contrasting RSA patterns. In tropical-tree communities there are fewer rare species in the local community than in the metacommunity, whereas the opposite pattern is found in coral reefs. Reference 6 reported log-series-like RSA distributions in local communities, and log-normal-like RSA distributions when a geographically widespread set of coral-reef communities was pooled to estimate the RSA distribution for the metacommunity. In contrast, local tropical-tree communities exhibit log-normal-like RSA distributions, which become more log-series-like at large landscape scales<sup>1–5</sup>. The log-series RSA distribution has a larger proportion of rare species than the log-normal. Here we consider two distinct types of community structure: first, a relatively small semi-isolated local community surrounded by a very large metacommunity acting as a source of immigrants, as in Hubbell's theory<sup>1</sup>, and, second, spatially isolated island communities whose assemblage acts as the metacommunity<sup>8,9</sup>. For the tropical forest, the timescale for species turnover in the metacommunity is very long compared to the characteristic timescale for immigration, leading to an effectively frozen metacommunity acting as a backdrop for immigration. In coral reefs, in contrast, each local community receives immigrants from all the surrounding semi-isolated local communities, within each of which the species abundances are not frozen in time. We present a simple unified theory for understanding the RSA patterns of tropical forests and coral reefs.

## Coral reefs

Consider a metacommunity consisting of many small semi-isolated local communities, each of which receives immigrants from other local communities. Because of the isolation of the local communities from each other, changes in the RSA distribution of the aggregated metacommunity may be assumed to occur more rapidly than immigration. One may make a simplifying assumption that the immigration parameter  $Y$  is species-independent, corresponding to immigration occurring not from a frozen metacommunity, as commonly assumed for a tropical forest, but from a time-averaged metacommunity in a species-symmetric manner. Using equations (12)

and (14) (see Box 1), the mean number of species with  $n$  individuals  $\langle \varphi_n \rangle$  is given by:

$$\langle \varphi_n \rangle = \theta \frac{x^n}{n!} \Gamma(n+Y) \quad (1)$$

where  $\theta = S/[(1-x)^{-Y} - 1]\Gamma(Y)$  is the Hubbell biodiversity number<sup>1</sup>,  $\Gamma$  is the gamma function,  $S$  is the number of observed species and  $x$  is the per capita birth-to-death-rate ratio.

Given the isolation of individual coral reefs under the island metacommunity model, the value of the immigration parameter  $Y$  is very small because the local communities are separated from each other by large distances. In such a situation, the RSA for the local communities resembles the Fisher log-series, and does not have an interior mode (at abundance  $n > 1$ ). Now let us gradually assemble the metacommunity RSA distribution by considering the joint RSA distributions of multiple local communities. First consider the joint RSA of two local communities A and B comprising the metacommunity. Consider a species that has  $n_A$  individuals in community A with probability  $P(n_A)$  and  $n_B$  individuals in community B with probability  $P(n_B)$ . The probability that the species has  $n$  individuals in A and B is (see Supplementary Materials):

$$P(n_A + n_B = n) = \sum_{n_A + n_B = n} P(n_A)P(n_B) \propto \frac{x^n}{n!} \Gamma(n+2Y) \quad (2)$$

The corresponding RSA has the same form as for the single community but with the effective immigration parameter  $2Y$ . Extending the calculation of the joint RSA distribution to more and more local communities, one arrives at the RSA of the metacommunity characterized by an effective immigration parameter  $LY$ , where  $L$  is the total number of local communities comprising the metacommunity. When  $L$  is large, the RSA distribution exhibits a clear, interior mode at abundance  $n > 1$ , and the rare species constitute a smaller fraction of all the species than in the local community. This mean field analysis does not take into account the actual spatial locations of the local reef communities. Figure 1 shows the fits of the island model to the coral-reef RSA data<sup>6,7</sup>. We can see from the fits that, on local scales (local communities and reefs), immigration is almost absent so

<sup>1</sup>Department of Physics, 104 Davey Laboratory, <sup>2</sup>Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. <sup>3</sup>Department of Ecology and Evolutionary Biology, The University of California, Los Angeles, California 90095, USA. <sup>4</sup>The Smithsonian Tropical Research Institute, Unit 0948, APO AA 34002, Panama. <sup>5</sup>Dipartimento di Fisica 'G. Galilei', Università di Padova CNISM and INFN, via Marzolo 8, 35131 Padova, Italy.

**Box 1 | General framework**

We neglect inter-species interactions after the community has reached a steady state, and consider the dynamics of the population of a single species. These dynamics are governed by generalized birth and death events (including speciation, immigration and emigration).  $b_{n,k}$  and  $d_{n,k}$  represent the probabilities of birth and death, respectively, in the  $k$ th species with  $n$  individuals with  $b_{-1,k} = d_{0,k} = 0$ .  $P_{n,k}(t)$  denotes the probability that the  $k$ th species contains  $n$  individuals at time  $t$ . In the simplest scenario, the time evolution of  $P_{n,k}(t)$  is regulated by the master equation<sup>21</sup>:

$$\frac{\partial P_{n,k}(t)}{\partial t} = P_{n-1,k}(t)b_{n-1,k} + P_{n+1,k}(t)d_{n+1,k} - P_{n,k}(t)(b_{n,k} + d_{n,k}) \quad (8)$$

which leads to the steady-state or equilibrium solution:

$$P_{n,k} = P_{0,k} \prod_{i=0}^{n-1} \frac{b_{i,k}}{d_{i+1,k}} \quad (9)$$

for  $n > 0$  and where  $P_{0,k}$  can be deduced from the normalization condition  $\sum_{n \geq 0} P_{n,k} = 1$ . One can show that the system is guaranteed to reach the stationary solution (9) in the infinite time limit<sup>22</sup>.

Let us consider a simple, ecologically meaningful form for the effective birth and death rates of the  $k$ th species:

$$b_{n,k} = b_k(n + Y_k) \quad (10)$$

and

$$d_{n,k} = d_k n \quad (11)$$

where  $b_k$  and  $d_k$  denote the per-capita density-independent birth and death rates and a non-zero  $Y_k$  could arise from either immigration or owing to intraspecific interactions such as those giving rise to density dependence<sup>3</sup>. We do not incorporate speciation explicitly into the model because it does not affect the functional form of the results (it can be incorporated into the immigration term at  $n = 0$  by adding a constant).

The steady-state solution of the master equation for  $P_k(n)$ , the probability that the  $k$ th species has  $n$  individuals, yields a negative binomial distribution<sup>23</sup>:

$$P_{n,k} = P_{0,k} \prod_{i=0}^{n-1} \frac{b_{i,k}}{d_{i+1,k}} = \frac{(1 - x_k)^{Y_k} x_k^n}{\Gamma(Y_k) n!} \Gamma(n + Y_k) \quad (12)$$

where  $x_k = b_k/d_k$ , the ratio of the per-capita birth rate to the per-capita death rate, controls the mean species abundance given by  $x_k Y_k / (1 - x_k)$ . Henceforth, we make the neutral ecological equivalence assumption that the per-capita birth and death rates are the same for all the species, that is,  $b_k = b$ ,  $d_k = d$  and  $x_k = x$ , so we can discard the subscript  $k$  for  $b$ ,  $d$  and  $x$ .

The number of species containing  $n$  individuals is given by:

$$\varphi_n = \sum_{k=1}^S I_{n,k} \quad (13)$$

where  $S$  is the total number of species that may potentially be present in the community and the indicator  $I_{n,k}$  is a random variable that takes the value 1 with probability  $P_{n,k}$  and 0 with probability  $(1 - P_{n,k})$ . Thus, the average number of species containing  $n$  individuals is given by:

$$\langle \varphi_n \rangle = \sum_{k=1}^S I_{n,k} = \sum_{k=1}^S P_{n,k} \quad (14)$$

and the variance is given by:

$$\sigma_{\varphi_n}^2 = \langle (\varphi_n - \langle \varphi_n \rangle)^2 \rangle = \left[ \sum_{k=1}^S (I_{n,k} - P_{n,k}) \right]^2 = \sum_{k=1}^S P_{n,k} (1 - P_{n,k}) \quad (15)$$

This result is based on the assumption that the fluctuations of  $I_{n,k}$  do not depend on the species label  $k$ , a consequence of the absence of interactions between species. The RSA relationship we seek to derive is the dependence of  $\langle \varphi_n \rangle$  on  $n$ . Another quantity of interest is the average number of species observed in the community:

$$\langle S_{\text{obs}} \rangle = S - \langle \varphi_0 \rangle = S - \sum_{k=1}^S (1 - x)^{Y_k} \quad (16)$$

For a sample of  $J$  individuals comprising  $\bar{S}$  species, the sampling multivariate probability distribution is a compound multinomial Dirichlet distribution<sup>19,24</sup>:

$$P(n_1, n_2, \dots, n_{\bar{S}} | J) = \frac{1}{P(J)} \prod_{k=1}^{\bar{S}} P_{n_k, k} \delta(J - n_1 - n_2 - \dots - n_{\bar{S}}) = \left( \frac{J + \sum_k Y_k - 1}{J} \right)^{-1} \prod_{k=1}^{\bar{S}} \binom{n_k + Y_k - 1}{n_k} \delta(J - n_1 - n_2 - \dots - n_{\bar{S}}) \quad (17)$$

where the normalization factor  $P(J)$  is the probability of observing a community of size  $J$ . Interestingly, the same formula also applies to community dynamics with a fixed total population obeying a zero-sum rule—that is, when the death of an individual is immediately followed by a birth or by the addition of an immigrant into the community. Refs 9 and 13 have derived the expression for the compound multinomial Dirichlet distribution<sup>19,24</sup> by using a master equation approach and have presented a sampling theory that considers both ecological and evolutionary times in a single sampling formula.

that the local RSA resembles the Fisher log-series distribution. Our theory explains how the RSA becomes log-normal-like on aggregating the local communities into one metacommunity.

We can estimate the species-similarity index between two local communities can be defined as:

$$\frac{2S_{A \cap B}}{S_A + S_B} 100\% \quad (3)$$

where  $S_A$  and  $S_B$  are the number of observed species in communities A and B, respectively, and  $S_{A \cap B}$  is the number of species that are present in both communities. Because immigration is almost absent for local communities, one can assume that for each local community the species are randomly drawn from the metacommunity species pool. With this assumption, the average similarity index becomes:

$$\frac{2S_A S_B}{S(S_A + S_B)} 100\% \quad (4)$$

where  $S$  is the total number of species in the metacommunity. Figure 2 shows the frequency distribution of the similarity index taken for all pairs of local communities (within a given region). The size of the species pool was chosen to be the same as the number of species in the regional metacommunity. Our results are in qualitative accord with those in ref. 7, which reported that the distribution of the similarity

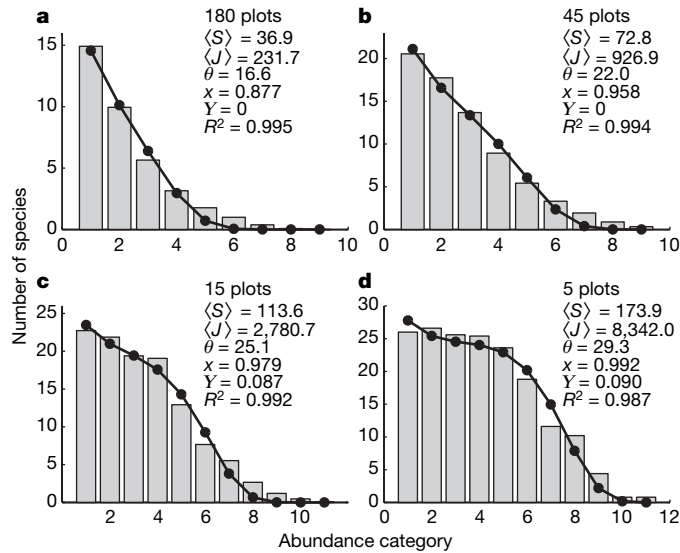
index among local coral reefs was quite different from the predictions of neutral theory valid for tropical forests. The average community similarity is low (about 35%) and has a large variance. A more refined model of the coral reefs would not only consider the biology of coral-reef reproduction but also the specific locations, environments and the degree of stationarity of local communities.

**Tropical forests**

We now consider the metacommunity model introduced in ref. 1, in which a local community is embedded within a surrounding metacommunity, which is a source of immigrants. The dynamics of the local community is governed by births, deaths and immigration, whereas the metacommunity is characterized by births, deaths and speciation. The characteristic rate of species turnover in the local community is much faster than in the metacommunity so that we may treat the metacommunity as a fixed backdrop for immigration processes. We consider the non-interacting species case; this means that we have to consider only one species at a time, greatly simplifying the theory (see Box 1).

For the metacommunity, we can introduce speciation with a rate  $v \ll 1$ , a very small probability for the creation of a new species. The  $v$  term does not contribute significantly when  $n \neq 0$ , but has a crucial role when  $n = 0$ . (Under the equivalence assumption of neutral theory, the species label of the new species is of no consequence.)





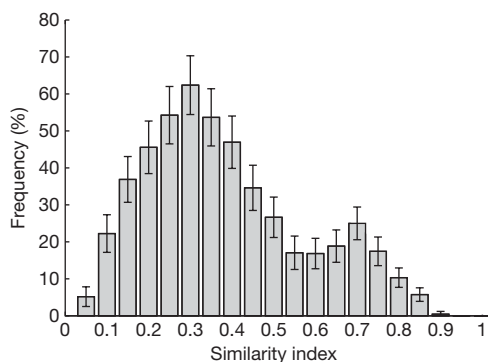
**Figure 1 | Relative species abundance of coral-reef communities.** We plot the fits of equation (1) (solid line) to the coral-reef species abundance data for the coral-reef local community (a), reef (b), metacommunity (c) and metacommunity (habitats pooled, d)<sup>6,7</sup>. The bars are observed numbers of species binned into log<sub>2</sub> abundance categories. The first histogram bar represents  $\langle \phi_1 \rangle$ , the second bar  $\langle \phi_2 \rangle + \langle \phi_3 \rangle$ , the third bar  $\langle \phi_4 \rangle + \langle \phi_5 \rangle + \langle \phi_6 \rangle + \langle \phi_7 \rangle$ , and so on. Graphs show the average RSA of: a, 180 local communities; b, 45 reef communities, each of which consists of 4 local communities; c, 15 metacommunities, each of which consists of 3 reef communities; d, metacommunities (habitats pooled), each of which consists of 5 metacommunities. Also shown are the values of the fitted parameters and the  $R^2$  values.

Operationally, this is mathematically equivalent to replacing  $Y_k$  (for the  $k$ th species) in equation (12) by  $v$ , which yields the logarithmic distribution  $P_n = vx^n/n + O(v^2)$ , where  $O(v^2)$  is a small term of the order of  $v^2$ .

Following ref. 2, one obtains an expression for the steady-state RSA distribution in the local community by setting  $Y_k = \tilde{m}p_k$ , where  $\tilde{m}$  is a measure of the immigration rate measured in units of the birth rate  $b$ , and  $p_k$  is the fraction of individuals in the surrounding metacommunity belonging to the  $k$ th species (see Box 2):

$$\langle \phi_n \rangle = \theta \frac{x^n}{n!} \int_0^\infty \frac{\Gamma(y+n)}{\Gamma(y+1)} e^{-\omega y} dy \quad (5)$$

where  $\omega = \theta/\tilde{m} - \ln(1-x)$  and  $\theta$  is the biodiversity parameter.



**Figure 2 | Similarity index for coral-reef communities.** We generated 10,000 random realizations of 60 samples with sizes equal to those of the coral-reef slopes<sup>7</sup> assuming that all the species have equal abundance in the metacommunity. The histogram shows the distribution of the similarity index predicted by the island model for the coral-reef data (the mean index is equal to 0.38 and the variance is 0.19) using equation (3). The distribution is qualitatively similar to that reported in ref. 7. Error bars represent standard deviation.

### Box 2 | Relative species abundance of tropical forests

The mean number of species with  $n$  individuals in a community can be written as:

$$\langle \phi_n \rangle = \sum_{k=1}^{S_M} P_{n,k} = S_M \int_0^\infty d\mu \hat{\rho}(\mu) P_{n,\mu} \quad (18)$$

where  $S_M$  is the number of species in the metacommunity. Here  $\hat{\rho}(\mu)d\mu$  is the continuous probability distribution of the mean populations of the species in the metacommunity and has the form of the familiar Fisher log-series (in a singularity-free description<sup>25,26</sup>):

$$\hat{\rho}(\mu)d\mu = \frac{1}{\Gamma(\varepsilon)\delta^\varepsilon} \exp(-\mu/\rho) \mu^{\varepsilon-1} d\mu \quad (19)$$

where  $\delta = x/(1-x)$  to match the first moment of the discrete and continuous distributions and:

$$P_{n,\mu} = \frac{(1-x)^{\tilde{m}\mu/J_M}}{\Gamma(\tilde{m}\mu/J_M)} \frac{x^n}{n!} \Gamma(n + \tilde{m}\mu/J_M) \quad (20)$$

where  $J_M$  is the total population of the metacommunity

( $J_M = \sum_{n=1}^\infty \theta x^n = \theta x/(1-x)$ ). Then, on substituting equations (19) and (20) into equation (18) and defining  $y = \tilde{m}\mu/J_M$ ,  $\varepsilon \rightarrow 0$  and  $\varepsilon S_M \rightarrow \theta$ , one obtains an analytical expression for the RSA of the local community:

$$\langle \phi_n \rangle = \theta \frac{x^n}{n!} \int_0^\infty dy \frac{\Gamma(n+y)}{\Gamma(1+y)} \exp\{-y[J_M/(\tilde{m}\delta) - \ln(1-x)]\} = \quad (21)$$

$$\theta \frac{x^n}{n!} \int_0^\infty \frac{\Gamma(n+y)}{\Gamma(1+y)} e^{-\omega y} dy \equiv \theta \frac{x^n}{n!} f(n, \omega)$$

where  $\omega = \theta/\tilde{m} - \ln(1-x)$  and the integral  $f(n, \omega)$  in the above equation can be calculated analytically using the following recurrence equations:  $f(1, \omega) = 1/\omega$ ,  $f(n+1, \omega) = nf(n, \omega) - \partial f(n, \omega)/\partial \omega$ . Equation (21) represents the average number of species with abundance  $n$  as a negative-binomial sampling from a metacommunity characterized by a log-series RSA<sup>10,13,27–29</sup>.

For the limiting case of no immigration ( $\tilde{m} \rightarrow 0$ ), the RSA of the local community approaches the Fisher log-series. For large immigration, for tropical forests with the Fisher log-series metacommunity ( $\tilde{m} \rightarrow \infty$ ), the probability distribution for each species is Poisson distributed with averages equal to those in the metacommunity so that one again obtains the Fisher log-series.

The expression equation (21) is much simpler than that presented in ref. 3:

$$\langle \phi_n \rangle = \theta \frac{J!}{n!(J-n)!} \frac{\Gamma(\gamma)}{\Gamma(J+\gamma)} \times \int_0^\gamma \frac{\Gamma(n+y)}{\Gamma(1+y)} \frac{\Gamma(J-n+\gamma-y)}{\Gamma(\gamma-y)} \exp(-y\theta/\gamma) dy \quad (22)$$

where  $\gamma = m(J-1)/(1-m)$ . The immigration parameter  $\tilde{m}$  in equation (21) is equal to  $Jm/(1-m)$  in ref. 3.

Using equation (5) we can readily obtain expressions for the average local community size  $\langle J \rangle$  and the average number of species  $S_{\text{obs}}$  in the local community:

$$\langle J \rangle = \sum_{n=1}^\infty n \langle \phi_n \rangle = \frac{\tilde{m}x}{1-x} \quad (6)$$

and

$$\langle S_{\text{obs}} \rangle = \sum_{n=1}^\infty \langle \phi_n \rangle = \theta \ln \left[ 1 - \frac{\tilde{m}}{\theta} \ln(1-x) \right] \quad (7)$$

Table 1 summarizes the results of fits of data from six large species-rich plant communities using equation (5) (see Fig. 3). These data have been analysed previously using different varieties of neutral

**Table 1 | Relative species abundance of tropical forests**

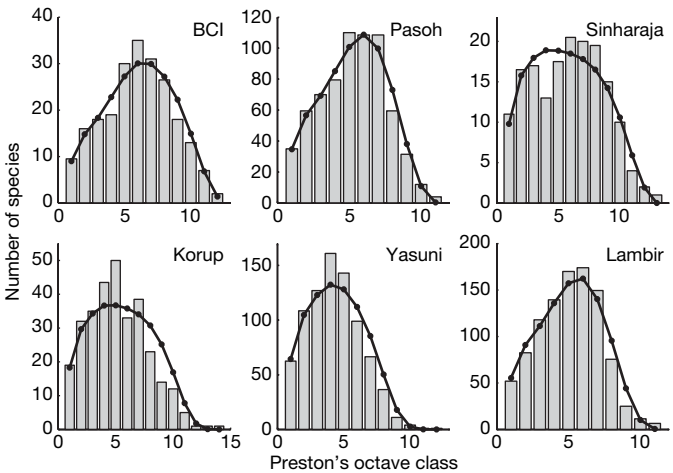
Plot	<i>S</i>	<i>J</i>	$\theta$	$\bar{m}$	<i>m</i>	<i>x</i>	<i>r</i> <sup>2</sup>
BCI, Panama	225	21,457	48	2,122	0.09	0.91	0.97
Yasuni, Ecuador	821	17,546	212	9,448	0.35	0.65	0.98
Pasoh, Malaysia	678	26,554	206	1,999	0.07	0.93	0.98
Korup, Cameroon	308	24,591	54	18,551	0.43	0.57	0.94
Lambir, Malaysia	1,004	33,175	305	3,281	0.09	0.91	0.99
Sinharaja, Sri Lanka	167	16,936	28	15,633	0.48	0.52	0.94

Estimates of the ecological parameters obtained from fitting equation (5) to the RSA data of six forests. *S*, the number of observed species; *J*, the local community size;  $\theta$ , the biodiversity parameter; *m*, the per capita immigration rate ( $\bar{m} = Jm/(1 - m)$ ); and *x*, the per capita birth-to-death-rate ratio. Substituting the above expression for  $\bar{m}$  into equation (6), one obtains  $m + x = 1$  (for the metacommunity  $x + v = 1$ ).

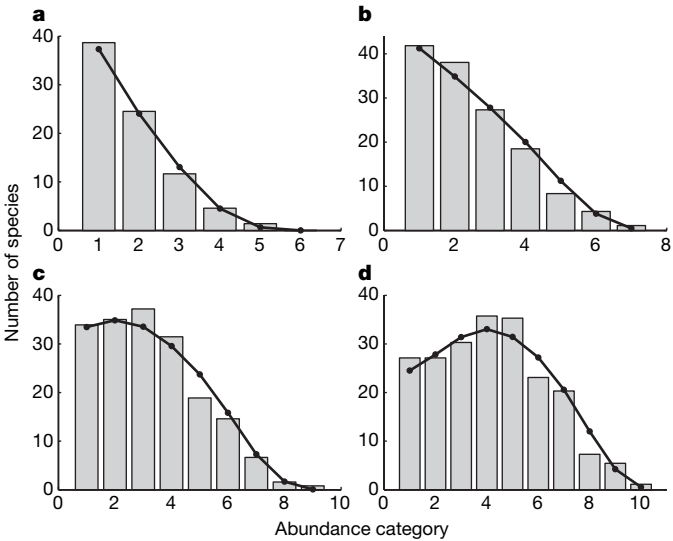
theory<sup>1–3,10</sup>. Our simple analytical approximations of the exact sampling theory<sup>8–14</sup> yield virtually indistinguishable fits to the data, and we obtain very similar values of the biological parameters to the ones derived previously. The exact theory applies both under the assumption of species independence in a fluctuating community and for species undergoing zero-sum dynamics<sup>9</sup>.

**Comparison between coral reefs and tropical forests**

To understand the qualitative difference between species composition in the coral-reef system and in tropical forests, we used the available relative tree species abundance data from the forest dynamics plot on Barro Colorado Island (BCI), Panama to generate random samples of the same size as in the coral-reef studies. Figure 4 shows the RSA for samples with 232, 927, 2,781 and 8,342 individuals (each histogram represents an RSA averaged over 100 samples). The abundance histograms are qualitatively similar to those for the coral-reef data. On fitting them with equation (5), the immigration parameter is very large for the small samples (for example, the per capita immigration rate  $m \approx 1$  for the sample of 232 individuals) and decreases with an increase in sample size. This can be explained as follows: for a small population, the number of immigrants exceeds the number of internal birth events. As the population size increases, immigration has a less important role and becomes negligible as the population goes to infinity (and thus forms a metacommunity). Thus, one would observe a Fisher log-series in two limiting cases: in the metacommunity in which there are no immigration events, and in the very small local community that has a high immigration rate from the metacommunity characterized by a Fisher log-series RSA. The sampled community of the coral reef represents a very small portion of the total reef population. However, the distinct similarity indices for the tropical forest and the coral-reef communities underscore a key difference in the two



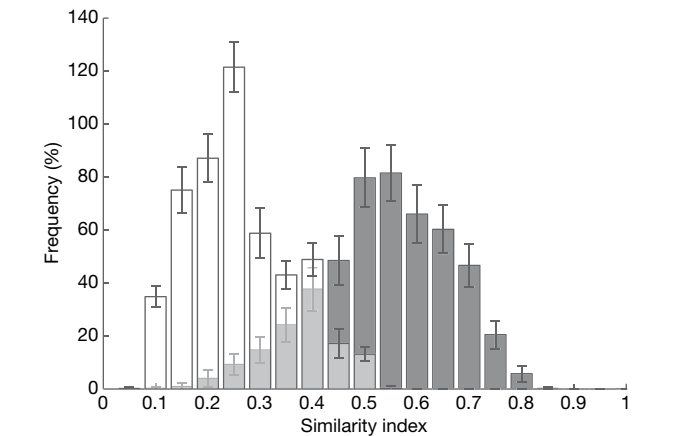
**Figure 3 | Relative species abundance of tropical forests.** We plot the fits of equation (5) (solid lines) to the tree species abundance data from the BCI, Pasoh, Sinharaja, Korup, Yasuni and Lambir plots (see Table 1). The frequency distributions are plotted using Preston's binning method<sup>1</sup>. The numbers on the x-axis represent Preston's octave classes.



**Figure 4 | Sampling of BCI tropical-forest data.** We plot the fits of equation (5) (solid line) to the RSA of samples chosen randomly from the BCI data. Each plot shows RSA for the same number of samples with the same abundances as in the coral-reef data. The RSAs are qualitatively similar to those in Fig. 1 (the values of *m* are equal to 1, 0.84, 0.46 and 0.19 for **a**, **b**, **c** and **d**, respectively).

cases. For tropical forests, the assumption of equation (4)—that every species is equally likely to occur in a given local community—leads to a similarity index distribution that is significantly different from the actual distribution given by equation (3) (see Fig. 5). In contrast, for the coral-reef community, the observed distribution is consistent with the prediction of equation (4) (see Fig. 2). These results suggest a high degree of isolation between local communities in the coral-reef ecosystem. Unlike in the tropical forest, the set of local communities in the coral-reef data are effectively uncorrelated samples from the metacommunity<sup>8,9</sup>.

Intraspecific interactions can be taken into account in a simplified manner by attributing density-dependent effective birth and death rates, which depend on the population of a species<sup>3</sup>. On retaining the equivalence of all species and within a parsimonious model, it was shown that the RSA behaves as  $\langle \phi_n \rangle \propto x^n/(n + c)$ , where a positive coefficient *c* describes a rare-species advantage<sup>3</sup>. The data presented in Figs 1 and 4 can also be fitted approximately with this model with effective values of *c* equal to 2.29, 1.26, 2.63, 0.87, 2.91 and 0.41 for tropical forests (BCI, Yasuni, Pasoh, Korup, Lambir and Sinharaja, respectively) and –0.24, 0.05, 0.30 and 0.42 for coral-reef communities (local, reef, metacommunity and metacommunity (habitats



**Figure 5 | Similarity index for the BCI plot.** We generated 10,000 random realizations of 60 samples with abundances equal to those of the coral reef slopes<sup>7</sup>. The dark and light histograms were calculated using equations (3) and (4), respectively. Error bars represent standard deviation.



pooled), respectively). Interestingly, the coral-reef local community is characterized by a negative  $c$  coefficient, which would correspond to a rare species disadvantage, analogous to the Allee effect<sup>15</sup>, in the effective birth-to-death-rates ratio.

## Discussion

We have presented a simple unified theoretical framework in which interspecific interactions are turned off, which yields analytical expressions for two distinct and very different types of metacommunities. One type<sup>1</sup>, a continuous source area from which immigrants to local communities are drawn, yields log-series RSA distributions on large scales and log-normal-like distributions with interior modes on local scales, as one consistently observes in tropical-forest communities. The other type<sup>8,9</sup>, a metacommunity consisting of an island archipelago of isolated communities, yields the opposite pattern: log-series-like RSA distributions in local communities, but interior-mode RSA distributions on large scales, as is observed in coral-reef communities.

In adopting this approach, we are not denying that species interact. There is a huge ecological literature on interspecific interactions, including competitive and predator–prey interactions, among others. What we are saying, however, is that we can make major theoretical progress in ecology by not considering species interactions at the outset. It is plausible that there is an underlying simplicity associated with an ecological community in the vicinity of its steady state. Although the effects of interactions could very well be a factor under non-equilibrium conditions, overt competition between a pair of species would be toned down in the steady state in favour of one or the other over local spatial and temporal scales. Indeed, what our theory shows is that a large fraction of the quantitative variation in patterns of RSA in tropical forests and coral reefs can be accounted for by our non-interacting species approach. Additional variation in patterns of RSA will undoubtedly be explained when theoretical ecologists incorporate species interactions, especially when we require the theory to be able to explain the abundances of named and identified species in ecological communities. This progression in the development of ecological theory has an analogy to the development of the theory of gases in physics<sup>16</sup>. First came the assumption that gases were ‘ideal’ and non-interacting, which then led to the ideal gas equation of state<sup>17</sup>. This does remarkably well, but deviations in the behaviour of gases from ideal behaviour led to refinements to the theory<sup>18</sup>. We look forward to analogous developments in theoretical ecology.

## METHODS SUMMARY

When confronting the theory with data, it is necessary to relax the zero-sum rule. For example, in the case of tropical forests, the sampling process entails counting all the trees within a fixed area, but there is always some variation in the total abundance of trees in a fixed sample area. This is not a problem at the stationary state, at which one can prove that the zero-sum rule can be relaxed<sup>9,19</sup> and the species can be treated as being independent of each other (see also Supplementary Material in ref. 3). Indeed, we have found that a maximum likelihood estimator (MLE) method<sup>19</sup> based on equation (17) on a coral-reef metacommunity system leads to a similar value of the parameter  $Y$  compared to the estimate based on the fitting of the average RSA with equation (5),  $Y_{MLE} = 0.058$  and  $Y_{RSA} = 0.087$ . We have checked the validity of the relaxation of the zero-sum rule by randomly generating 1,000 samples and calculating the exact likelihood function for them. As expected, the coral-reef data are indistinguishable from the random pseudo-samples. For the fitting procedure, we assumed that the census data on species composition represents the average RSA. In all figures, we estimated the model parameters by a least-squares fit, minimizing the sum of squared residuals<sup>20</sup>.

Received 8 June; accepted 24 August 2007.

- Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, Princeton, 2001).
- Volkov, I., Banavar, J. R., Hubbell, S. P. & Maritan, A. Neutral theory and relative species abundance in ecology. *Nature* **424**, 1035–1037 (2003).

- Volkov, I., Banavar, J. R., He, F., Hubbell, S. P. & Maritan, A. Density and frequency dependence explains tree species abundance and diversity in tropical forests. *Nature* **438**, 658–661 (2005).
- Latimer, A. M., Silander, J. A. Jr & Cowling, R. M. Neutral ecological theory reveals isolation and rapid speciation in a biodiversity hot spot. *Science* **309**, 1722–1725 (2005).
- Etienne, R. S., Latimer, A. M., Silander, J. A. Jr & Cowling, R. M. Comment on “Neutral ecological theory reveals isolation and rapid speciation in a biodiversity hot spot”. *Science* **311**, 610 (2006).
- Connolly, S. R., Hughes, T. P., Bellwood, D. R. & Karlson, R. H. Community structure of corals and reef fishes at multiple scales. *Science* **309**, 1363–1365 (2005).
- Dornelas, M., Connolly, S. R. & Hughes, T. P. Coral reef diversity refutes the neutral theory of biodiversity. *Nature* **440**, 80–82 (2006).
- Alonso, D. & Pascual, M. Comment on “A keystone mutualism drives pattern in a power function”. *Science* **313**, 1739 (2006).
- Etienne, R. S., Alonso, D. & McKane, A. J. The zero-sum assumption in neutral biodiversity theory. *J. Theor. Biol.* **248**, 522–536 (2007).
- Alonso, D. & McKane, A. J. Sampling Hubbell’s neutral theory of biodiversity. *Ecol. Lett.* **7**, 901–910 (2004).
- Etienne, R. S. & Olff, H. A novel genealogical approach to neutral biodiversity theory. *Ecol. Lett.* **7**, 170–175 (2004).
- Etienne, R. S. A new sampling formula for neutral biodiversity. *Ecol. Lett.* **8**, 253–260 (2005).
- Etienne, R. S. & Alonso, D. A dispersal-limited sampling theory for species and alleles. *Ecol. Lett.* **8**, 1147–1156 (2005); erratum **9**, 500 (2006).
- Etienne, R. S. & Alonso, D. Neutral community theory: how stochasticity and dispersal-limitation can explain species coexistence. *J. Stat. Phys.* **128**, 485–510 (2006).
- Courchamp, F., Clutton-Brock, T. & Grenfell, B. Inverse density dependence and the Allee effect. *Trends Ecol. Evol.* **14**, 405–410 (1999).
- Harte, J. Tail of death and resurrection. *Nature* **424**, 1006–1007 (2003).
- Maxwell, J. C. *The Scientific Papers of James Clerk Maxwell* Vol. 1 (Dover, New York, 2003).
- van der Waals, J. D. *On the Continuity of the Gaseous and Liquid States* (Dover, New York, 2004).
- Rannala, B. The sampling theory of neutral alleles in an island population of fluctuating size. *Theor. Popul. Biol.* **50**, 91–104 (1996).
- Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge Univ. Press, Cambridge, 1993).
- Feller, W. *An Introduction to Probability Theory and Its Applications* Vol. 1 (Wiley & Sons, Hoboken, 1968).
- Van Kampen, N. G. *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 2001).
- Kendall, D. G. Stochastic processes and population growth. *J. Roy. Statist. Soc. B* **11**, 230–282 (1949).
- Mosimann, J. E. On the compound multinomial distribution, the multivariate distribution, and correlations among proportions. *Biometrika* **49**, 65–82 (1962).
- Fisher, R. A., Corbet, A. S. & Williams, C. B. The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* **12**, 42–58 (1943).
- Rao, C. R. *Statistical Ecology* Vol. 1, *Spatial Patterns and Statistical Distributions* 131–142 (Penn. State Univ. Press, University Park, Pennsylvania, 1971).
- Pielou, E. C. *An Introduction to Mathematical Ecology* (Wiley, New York, 1969).
- Bulmer, M. G. On fitting the Poisson lognormal distribution to species-abundance data. *Biometrics* **30**, 101–110 (1974).
- Dewdney, A. K. A general theory of the sampling process with applications to the veil line. *Theor. Popul. Biol.* **54**, 294–302 (1998).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank D. Alonso for bringing some key references to our attention and for spending an enormous amount of time in helping improve our paper. We thank D. Thomas, G. Chuyong and D. Kenfack for the data from Korup National Park, Cameroon; R. Valencia, R. Foster and R. Condit for the data from Yasuni National Park, Ecuador; S. Davies, S. Tan, J. LaFrankie and P. Ashton for the data from Lambir Hills National Park, Sarawak; N. Supardi, P. Ashton and J. LaFrankie for the data from Pasoh Forest Reserve, Peninsular Malaysia; and Hubbell’s collaborators on the Barro Colorado Island plot, R. Foster and R. Condit. We also thank S. Davies for directing and coordinating the global programs of the Center for Tropical Forest Science (CTFS), which manages the plots, S. Loo for data management, and I. Rubinoff, Director of the Smithsonian Tropical Research Institute, the host institution of CTFS. The fieldwork has also received long-term support from the John D. and Catherine T. MacArthur Foundation, the Mellon Foundation, Earthwatch, Frank Levinson and the Celera Foundation, and other private foundations and individual donors. We thank S. Connolly, M. Dornelas and T. Hughes for sending us the coral-reef data. This work was supported by COFIN 2005 and by the NSF.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to J.R.B. (banavar@psu.edu) or A.M. (maritan@pd.infn.it).

## ARTICLES

# The origin of spontaneous activity in the developing auditory system

Nicolas X. Tritsch<sup>1</sup>, Eunyoung Yi<sup>2</sup>, Jonathan E. Gale<sup>3</sup>, Elisabeth Glowatzki<sup>1,2</sup> & Dwight E. Bergles<sup>1,2</sup>

**Spontaneous activity in the developing auditory system is required for neuronal survival as well as the refinement and maintenance of tonotopic maps in the brain. However, the mechanisms responsible for initiating auditory nerve firing in the absence of sound have not been determined. Here we show that supporting cells in the developing rat cochlea spontaneously release ATP, which causes nearby inner hair cells to depolarize and release glutamate, triggering discrete bursts of action potentials in primary auditory neurons. This endogenous, ATP-mediated signalling synchronizes the output of neighbouring inner hair cells, which may help refine tonotopic maps in the brain. Spontaneous ATP-dependent signalling rapidly subsides after the onset of hearing, thereby preventing this experience-independent activity from interfering with accurate encoding of sound. These data indicate that supporting cells in the organ of Corti initiate electrical activity in auditory nerves before hearing, pointing to an essential role for peripheral, non-sensory cells in the development of central auditory pathways.**

Auditory perception depends on the precise conversion of sound-induced vibrations of the basilar membrane into graded release of transmitter from inner hair cells (IHCs), which provide the main excitatory input to auditory nerve fibres. The immaturity of the middle and inner ear of newborn rats prevents the detection of airborne sounds before the ‘onset of hearing’—the age range over which neonates first display sound-evoked neural responses (postnatal day (P) 11–13)<sup>1,2</sup>. Nevertheless, IHCs can release glutamate in a  $\text{Ca}^{2+}$ -dependent manner<sup>3</sup> and auditory nerves fire discrete bursts of action potentials before the onset of hearing<sup>4–7</sup>, suggesting that spiral ganglion neurons (SGNs) are subjected to periodic excitation. This spontaneous activity is abolished after cochlea removal or application of the sodium channel blocker tetrodotoxin (TTX) to the oval window<sup>8</sup>, indicating that the trigger for this activity resides within the cochlea. However, the mechanisms responsible for initiating auditory nerve firing in the absence of sound have not been determined.

## Spontaneous ATP release in the developing cochlea

The developing cochlea of mammals contains a transient structure of unknown function termed Kölliker’s organ (or greater epithelial ridge), consisting of a wide expanse of columnar-shaped supporting cells<sup>9–11</sup> (Fig. 1a). In cochlear turns acutely isolated from rats before the onset of hearing, we observed spontaneous inward currents in whole-cell voltage-clamp recordings from these ‘inner’ supporting cells. These events occurred at a frequency of  $0.20 \pm 0.01$  Hz, varied widely in amplitude (mean =  $-254 \pm 20$  pA, coefficient of variance =  $1.48 \pm 0.03$ ) and exhibited remarkably slow kinetics (mean rise time,  $t_{\text{rise}} = 1,546 \pm 51$  ms,  $n = 23$ ; Fig. 1b). Despite their low input resistance ( $7.4 \pm 0.4$  M $\Omega$ ,  $n = 16$ ), inner supporting cells were depolarized by as much as 37 mV by these spontaneous inward currents (Fig. 1c). When a field electrode was inserted within Kölliker’s organ, spontaneous extracellular potentials up to 5 mV in amplitude (mean =  $0.21 \pm 0.01$  mV,  $n = 66$ ) were observed; these responses exhibited slow kinetics ( $t_{\text{rise}} = 1,098 \pm 47$  ms) and occurred with a frequency similar to that of spontaneous inward currents ( $0.19 \pm 0.01$  Hz; Fig.

1d). In simultaneous recordings, 85% of spontaneous extracellular potentials were coincident with spontaneous inward currents in nearby supporting cells ( $n = 447$  field potentials in seven pairs; Supplementary Fig. 1), suggesting that extracellular potentials result from the spontaneous currents. These results indicate that there is widespread intrinsic activity in the developing cochlea, which can be detected as inward currents in supporting cells and local field potentials within Kölliker’s organ.

To determine the mechanisms responsible for this spontaneous activity, we applied pharmacological inhibitors of receptors known to be expressed in the developing cochlea<sup>12</sup>. Field potentials were not altered by the  $\alpha_9\alpha_{10}$  acetylcholine receptor antagonist strychnine ( $n = 5$ ,  $P > 0.7$ ), the muscarinic receptor antagonist atropine ( $n = 4$ ,  $P > 0.7$ ), or the nicotinic receptor antagonist D-tubocurarine (D-TC,  $n = 4$ ,  $P > 0.4$ ), indicating that this spontaneous activity was not initiated by cholinergic efferents<sup>13</sup>. These events also did not require  $\text{Ca}^{2+}$ -dependent release of glutamate from IHCs or neuronal firing, as they were not affected by cadmium ( $n = 4$ ,  $P > 0.1$ ) or nifedipine ( $n = 4$ ,  $P > 0.9$ ), which inhibit IHC  $\text{Ca}^{2+}$  currents<sup>14</sup>, by the glutamate receptor antagonists 2,3-dihydroxy-6-nitro-7-sulphamoyl-benzo(f) quinoxaline (NBQX,  $n = 4$ ,  $P > 0.8$ ) and *R,S*-3-(2-carboxypiperazin-4-yl)propyl-1-phosphonic acid (*R,S*-CPP,  $n = 3$ ,  $P > 0.5$ ), or by TTX ( $n = 3$ ,  $P > 0.6$ ) (Fig. 1e and Supplementary Fig. 2). However, spontaneous extracellular potentials were significantly inhibited by the P2 purinergic receptor antagonists pyridoxal-phosphate-6-azophenyl-2’,4’-disulphonate (PPADS,  $n = 5$ ,  $P < 0.0001$ ) and suramin ( $n = 6$ ,  $P < 0.0005$ ), but not the adenosine receptor antagonist 1,3-dipropyl-8-cyclopentylxanthine (DPCPX,  $n = 4$ ,  $P > 0.4$ ), suggesting that this spontaneous cochlear activity was initiated by extracellular ATP (Fig. 1d, e and Supplementary Fig. 2). Consistent with this hypothesis, both the frequency and amplitude of spontaneous extracellular potentials were significantly decreased by the ATP-hydrolysing enzyme apyrase ( $n = 5$ ,  $P < 0.005$ ). Moreover, in whole-cell recordings from inner supporting cells, application of ATP, but not adenosine, induced large inward currents ( $4.1 \pm 0.8$  nA,  $n = 7$ ; Supplementary Fig. 3), indicating that ATP is sufficient to elicit this activity.

<sup>1</sup>The Solomon H. Snyder Department of Neuroscience, <sup>2</sup>The Center for Hearing and Balance, Department of Otolaryngology—Head and Neck Surgery, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, USA. <sup>3</sup>UCL Ear Institute and Department of Physiology, University College London, London WC1X 8EE, UK.



Extracellular ATP exerts modulatory effects by binding to ionotropic (P2X) and metabotropic (P2Y) purinergic receptors in target tissues<sup>15</sup>, and members of both receptor classes are expressed in the developing organ of Corti<sup>16</sup>. Indeed, ATP-evoked currents in inner supporting cells exhibited prominent inward rectification and reversed near 0 mV ( $n = 5$ ; Supplementary Fig. 3), properties consistent with recombinant P2X receptors<sup>17</sup>. Large inward currents were also elicited in supporting cells by the P2Y receptor agonist UTP ( $n = 6$ ; Supplementary Fig. 3), suggesting that both P2X and P2Y receptors contribute to the ATP-dependent activity observed in Kölliker's organ.

Studies of epithelial cells in the periphery and glia cells in the central nervous system indicate that ATP can be released through large transmembrane pores such as P2X<sub>7</sub> receptors or unpaired connexons, termed hemichannels<sup>18</sup>. Cochlear supporting cells are extensively coupled through gap junctions, and connexin 26, which is

highly expressed by these cells<sup>19</sup>, has been shown to allow the passage of large anions, such as ATP<sup>20</sup>. ATP-dependent spontaneous activity in Kölliker's organ was blocked by the gap junction inhibitors octanol ( $n = 6$ ,  $P < 0.0001$ ) and carbenoxolone (CBX,  $n = 4$ ,  $P < 0.0001$ ). In addition, the frequency of spontaneous extracellular potentials was markedly increased by exposure to  $\text{Ca}^{2+}$ -free solution, a manipulation that opens hemichannels<sup>18</sup>, and this increase in activity was inhibited by PPADS ( $n = 4$ ,  $P < 0.018$ ; Fig. 1e). Conversely, spontaneous activity was not affected by the P2X<sub>7</sub> receptor antagonist brilliant blue G ( $n = 5$ ,  $P > 0.3$ ; Supplementary Fig. 2). These results suggest that gap junctions, and perhaps unpaired connexon hemichannels, are required for ATP release in the developing cochlea.

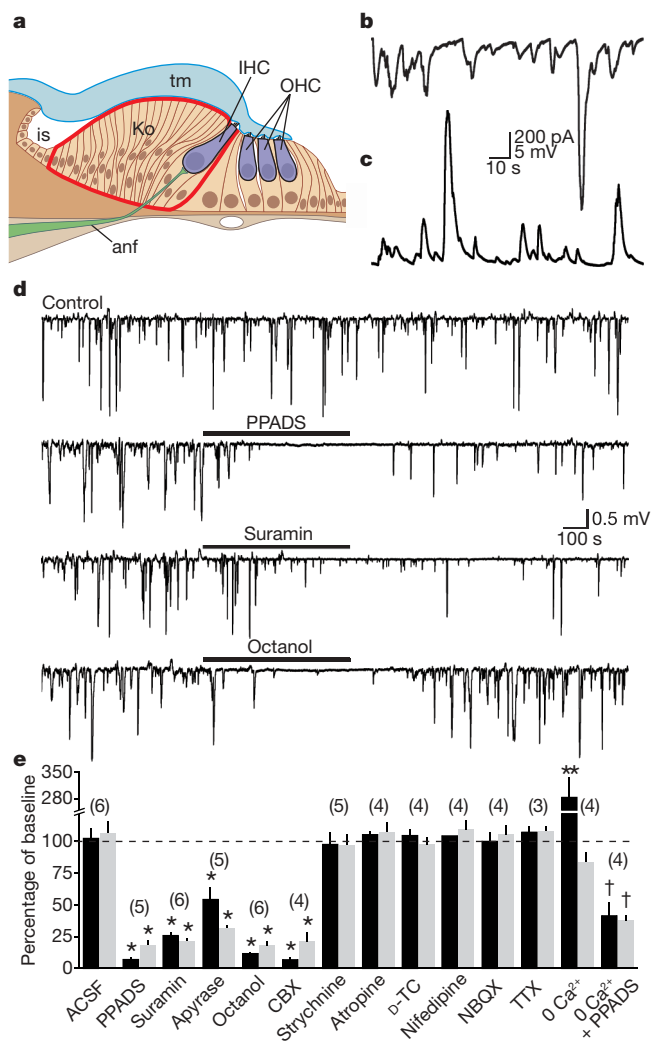
### ATP induces changes in cell shape

While imaging acutely isolated organs of Corti, periodic changes in light scattering were detected around groups of supporting cells within Kölliker's organ (Fig. 2a–c; Supplementary Video 1a–c). These spontaneous optical changes occurred randomly throughout Kölliker's organ at a frequency of  $0.034 \pm 0.003$  Hz ( $n = 15$ ), and were also observed along the thin processes of phalangeal cells, supporting cells that separate adjacent IHCs (Fig. 2d). In contrast, spontaneous optical changes were never observed near outer hair cells or in the outer sulcus. High-magnification imaging revealed that these local changes in transmittance resulted from an increase in extracellular space after crenation of inner supporting cells (Supplementary Fig. 4). As with the electrical activity recorded in this region, spontaneous optical changes were inhibited  $94 \pm 4\%$  by PPADS ( $n = 8$ ,  $P < 0.0001$ ),  $94 \pm 2\%$  by suramin ( $n = 7$ ,  $P < 0.0001$ ) and  $91 \pm 4\%$  by octanol ( $n = 7$ ,  $P < 0.0001$ ; Fig. 2e), and could be elicited by focal application of ATP or UTP to Kölliker's organ (Supplementary Fig. 3). In experiments where imaging was performed simultaneously with whole-cell recording from inner supporting cells, 93% of spontaneous optical changes were found to be associated with spontaneous inward currents ( $n = 380$  optical events in 12 cochleae; Fig. 2f), indicating that spontaneous optical changes result from the same endogenous ATP release events that mediate inward currents in supporting cells. Thus, intrinsic optical changes provide a non-invasive method for monitoring the spatial and temporal dynamics of ATP release in the organ of Corti.

Time-lapse imaging revealed that these optical changes propagated in a wave-like manner among supporting cells at 5 to  $15 \mu\text{m s}^{-1}$  (Fig. 2a–c; Supplementary Video 1a–c), rates comparable to the propagation of ATP-dependent  $\text{Ca}^{2+}$  waves among astrocytes<sup>21</sup>. In cochlear explant cultures loaded with the  $\text{Ca}^{2+}$  indicator dye fura-2, we observed spontaneous elevations of intracellular  $\text{Ca}^{2+}$  within the supporting cells of Kölliker's organ that propagated as waves ( $n = 7$  cochleae; Fig. 2g). Simultaneous differential interference contrast (DIC) and fluorescence imaging revealed that these  $\text{Ca}^{2+}$  transients either preceded or occurred coincident with spontaneous optical changes ( $n = 23$ ; Fig. 2g, h), indicating that crenation follows the rise in intracellular  $\text{Ca}^{2+}$  concentration ( $[\text{Ca}^{2+}]_i$ ). The frequency of spontaneous  $\text{Ca}^{2+}$  transients, like spontaneous optical changes, was decreased  $80 \pm 9\%$  by suramin ( $n = 4$ ,  $P < 0.003$ ). These results suggest that extracellular ATP activates purinergic autoreceptors on inner supporting cells, leading to a rise in  $[\text{Ca}^{2+}]_i$  and crenation. Supporting cells lateral to IHCs (for example, Deiters' cells, Hensen cells) also express purinergic receptors, and focal application of ATP to these cells triggered  $\text{Ca}^{2+}$  waves that propagated radially<sup>22</sup> (Supplementary Fig. 5). However, spontaneous  $\text{Ca}^{2+}$  waves were never observed in these cells ( $n = 16$  cochleae). Together, these results suggest that ATP is primarily released from supporting cells within Kölliker's organ.

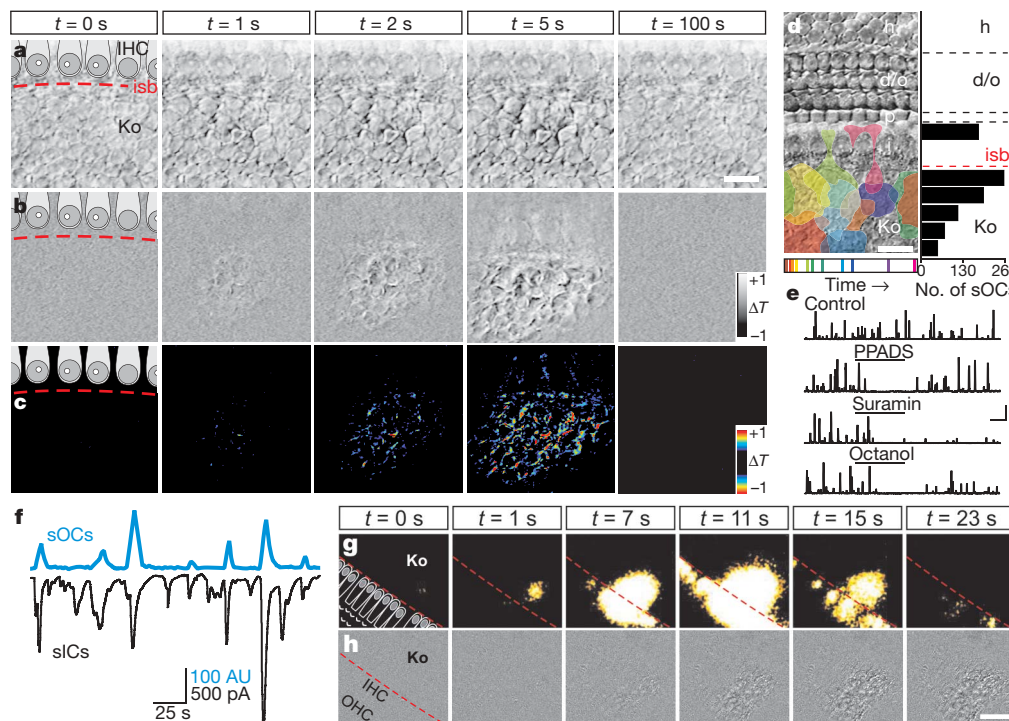
### ATP excites hair cells and afferent nerve fibres

Previous studies have demonstrated that purinergic receptors are also expressed by IHCs and SGNs<sup>23,24</sup>. As most ATP-dependent spontaneous optical changes originated close to IHCs (Fig. 2d), we



**Figure 1 | Spontaneous purinergic signalling in the developing cochlea.**

**a**, Diagram of a P7 rat organ of Corti in cross-section. anf, auditory nerve fibre; IHC, inner hair cell; is, inner sulcus; Ko, Kölliker's organ (outlined in red); OHC, outer hair cell; tm, tectorial membrane (adapted from refs 9–11). **b**, **c**, Spontaneous activity recorded in voltage-clamp (**b**) and current-clamp (**c**, resting membrane potential ( $V_m$ ) =  $-86$  mV) from a P7 supporting cell. **d**, Spontaneous extracellular potentials recorded from Kölliker's organ showing the effects of PPADS ( $50 \mu\text{M}$ ), suramin ( $150 \mu\text{M}$ ) and octanol ( $1 \text{ mM}$ ). **e**, Histogram showing the effects of antagonists on frequency (black) and amplitude (grey) of spontaneous extracellular potentials. Asterisk,  $P < 0.0005$  versus ACSF; double asterisk,  $P < 0.05$  versus ACSF; dagger,  $P < 0.02$  versus  $0 \text{ Ca}^{2+}$  ACSF. Data in **e** represent mean  $\pm$  s.e.m. The number of experiments is indicated in parentheses.



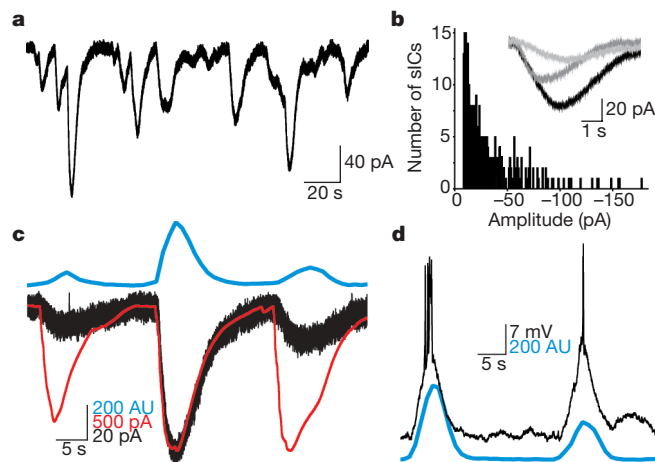
**Figure 2 | ATP release elicits optical changes and intercellular  $\text{Ca}^{2+}$  waves in Kölliker's organ.** **a–c**, Spontaneous optical changes within Kölliker's organ. **a**, Sequence of DIC images at indicated times ( $t$ ). Dashed line demarcates the inner spiral bundle. **b**, Difference images after subtraction (see Methods). **c**, Pseudocoloured difference images after thresholding. **d**, Left: location and furthest extent of spontaneous optical changes (sOCs) during 300 s (time of occurrence is shown at the bottom). Right: distribution of origins of spontaneous optical changes in 10  $\mu\text{m}$  bins ( $n = 879$  events in 21 preparations). **e**, Plots of normalized transmittance changes ( $\Delta T$ ) over time;

spontaneous optical changes are shown as upward deflections. Scale: 400 AU (arbitrary units), 200 s. **f**, Simultaneous recording of supporting cell spontaneous inward currents (black) and spontaneous optical changes (blue) from Kölliker's organ. sICs, spontaneous inward currents. **g, h**, Simultaneous time-lapse imaging of  $[\text{Ca}^{2+}]_i$  (**g**) and intrinsic optical changes (difference images, **h**). d/o, Deiters' and outer hair cells; h, Hensen cells; i/IHC, inner hair cells; isb, inner spiral bundle; Ko, Kölliker's organ; OHC, outer hair cells; p, pillar cells. Scale bars for **a–c**, 15  $\mu\text{m}$ ; **d**, 20  $\mu\text{m}$ , **g, h**, 30  $\mu\text{m}$ .

investigated whether the periodic release of ATP also affected IHCs. When care was taken to maintain the integrity of the tissue, we detected spontaneous inward currents in all IHCs ( $n = 35$ ), which occurred at a frequency of  $0.053 \pm 0.003$  Hz (Fig. 3a) and ranged in amplitude from  $-8$  to  $-180$  pA (mean =  $-23.0 \pm 0.9$  pA, coefficient of variance =  $0.71 \pm 0.03$ ; Fig. 3b). To determine the relationship between these slow currents ( $t_{\text{rise}} = 2,222 \pm 80$  ms) and the activity recorded from inner supporting cells, we recorded simultaneously from IHCs and nearby supporting cells while imaging intrinsic optical changes within Kölliker's organ. As shown in Fig. 3c, spontaneous currents in IHCs were coincident with both spontaneous inward currents in supporting cells and spontaneous optical changes, suggesting that IHCs respond to the same ATP release events. Consistent with this hypothesis, PPADS and suramin decreased the frequency of spontaneous IHC currents by  $92 \pm 4\%$  and  $89 \pm 4\%$ , respectively ( $n = 5$ ,  $P < 0.0001$ ; Supplementary Fig. 6). The membrane conductance of IHCs increased by as much as 28% during these events, and the magnitude of the conductance change was proportional to the size of the current ( $r = 0.82$ ,  $n = 97$  events in four IHCs), consistent with the activation of ionotropic P2 receptors. Current-clamp recordings from IHCs revealed that these inward currents depolarized IHCs by as much as 28 mV and were capable of triggering bursts of  $\text{Ca}^{2+}$  action potentials ( $n = 18$ ; Fig. 3d). In the presence of PPADS, the resting membrane potential of IHCs remained stable, except for small efferent-induced hyperpolarizations<sup>13</sup>. Together, these results suggest that the primary excitatory drive to IHCs at this age comes from ATP released by supporting cells in Kölliker's organ.

Previous studies indicate that  $\text{Ca}^{2+}$  spikes in neonatal IHCs induce bursts of excitatory postsynaptic currents (EPSCs) in afferent dendrites<sup>3</sup>. To address whether ATP-induced depolarization of IHCs

triggers  $\text{Ca}^{2+}$ -dependent release of glutamate at IHC-afferent synapses, we made whole-cell recordings from the dendritic enlargements of afferent neurons at the base of IHCs and measured their



**Figure 3 | Supporting-cell-derived ATP depolarizes IHCs.** **a**, Spontaneous inward currents recorded from a P10 IHC. **b**, Amplitude distribution of spontaneous inward currents (sICs) recorded from IHC in **a**. Inset: three representative spontaneous inward currents. **c**, IHC spontaneous inward currents (black trace) were coincident with spontaneous inward currents recorded from a nearby inner supporting cell (red trace) and spontaneous optical changes (blue trace) imaged in the adjacent region of Kölliker's organ. **d**, Spontaneous depolarizations recorded from a P9 IHC (black trace;  $V_m = -72$  mV) triggered  $\text{Ca}^{2+}$  spikes that were coincident with spontaneous optical changes (blue trace) in the adjacent region of Kölliker's organ.



response to exogenous ATP. In 7 out of 8 dendritic recordings, ATP application elicited a burst of EPSCs (Fig. 4a), presumably resulting from depolarization of and subsequent glutamate release from the presynaptic IHC. ATP also induced a slow inward current ( $-63 \pm 18$  pA,  $n = 8$ ) that reversed near 0 mV, suggesting that ATP has a direct excitatory effect on these afferents. In current-clamp recordings, ATP triggered a burst of excitatory postsynaptic potentials in afferent dendrites ( $n = 4$  of 5 recordings; Fig. 4b) that was blocked by NBQX ( $n = 3$ ), indicating that the ATP-induced increase in activity requires glutamate release from IHCs.

To determine whether spontaneous release of ATP by inner supporting cells is sufficient to elicit transmitter release from IHCs, we monitored spontaneous optical changes in the region where

dendritic recordings were made as an index of endogenous purinergic signalling. In four continuous recordings from afferent dendrites, discrete bursts of EPSCs occurred, separated by long periods with little synaptic activity (Fig. 4c). These bursts of EPSCs occurred at a frequency of  $0.020 \pm 0.005$  Hz and lasted  $6.9 \pm 0.8$  s ( $n = 46$ ). Each burst was superimposed on a small increase in holding current ( $-7 \pm 1$  pA), similar to the responses elicited by exogenous ATP. Furthermore, 87% of EPSC bursts were coincident with spontaneous optical changes, and 71% of spontaneous optical changes originating within  $50 \mu\text{m}$  of the terminals correlated with bursts, indicating that most ATP release events induce burst activity in nearby afferents.

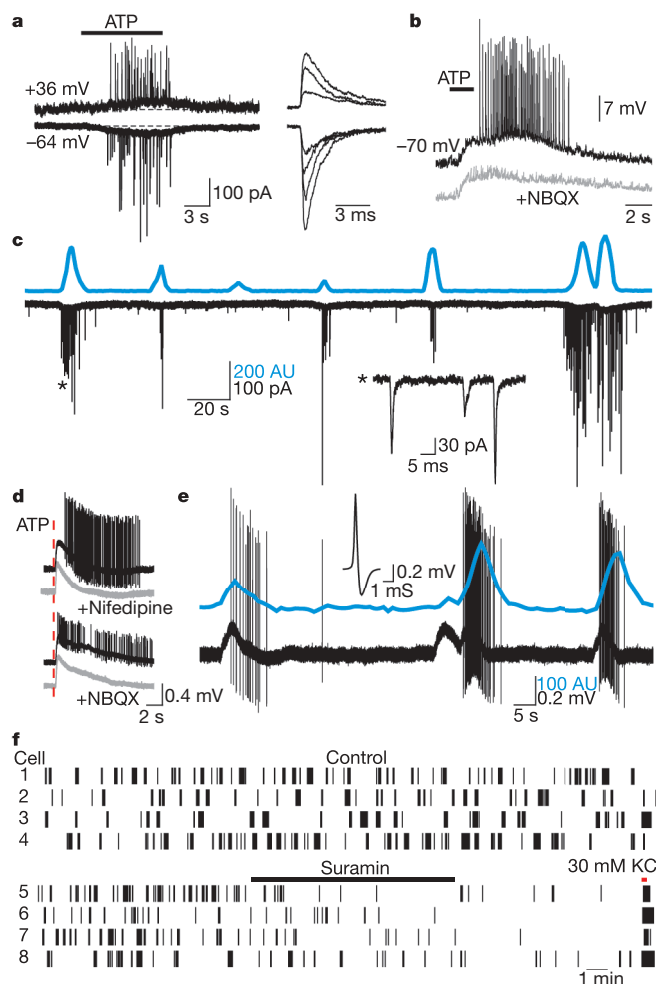
To test whether extracellular ATP is responsible for initiating bursts of action potentials in auditory nerve fibres, we made loose-patch recordings from the somata of SGNs in cochlear whole mounts. When ATP was applied near the IHC presynaptic to the neuron being monitored, a train of action potentials was elicited ( $n = 16$ ; Fig. 4d). This ATP-dependent increase in firing rate was blocked by nifedipine ( $n = 3$ ) and NBQX ( $n = 7$ ), indicating that it required  $\text{Ca}^{2+}$ -dependent glutamate release from IHCs. In continuous recordings from SGNs, discrete bursts of action potentials were observed at a frequency of  $0.014 \pm 0.001$  Hz ( $n = 11$ ; Fig. 4e), similar to the pattern of spontaneous EPSCs recorded from the dendrites of these neurons. Moreover, 90% of these bursts were coincident with spontaneous optical changes that occurred in the vicinity of the presynaptic IHC ( $n = 244$  bursts). If this periodic firing is caused by spontaneous release of ATP, then it should be inhibited by purinergic receptor antagonists. Indeed, the frequency of spontaneous action potential bursts in SGNs was decreased  $79 \pm 8\%$  by suramin ( $n = 11$ ,  $P = 0.0001$ ; Fig. 4f). These results indicate that bursts of action potentials in developing auditory nerve fibres are initiated by the periodic release of ATP from supporting cells in Kölliker's organ.

### Spontaneous ATP signalling ceases at hearing onset

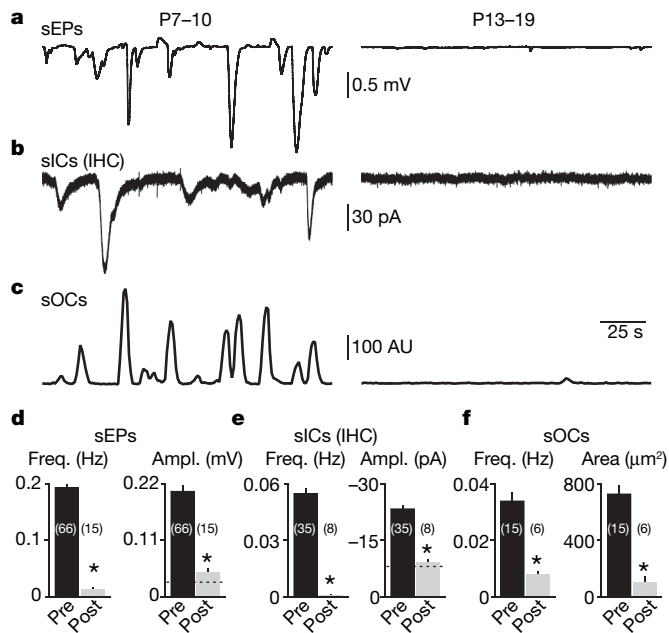
The pattern of auditory nerve fibre activity shifts from bursting to continuous discharge after the onset of hearing<sup>25,26</sup>, suggesting that episodic activation of IHCs by ATP may decrease with age. To determine whether spontaneous purinergic signalling in the cochlea diminishes with development, we compared intrinsic ATP-mediated responses in organs of Corti isolated from rats before (P7–10) and after (P13–19) the onset of hearing<sup>1,2</sup>. We found that spontaneous optical changes, spontaneous extracellular potentials and spontaneous inward currents in IHCs were almost completely absent after the onset of hearing (Fig. 5). Thus, endogenous purinergic signalling in the cochlea follows the developmental change in the pattern of intrinsic activity observed *in vivo*. This rapid developmental down-regulation may help to ensure faithful representation of experience-dependent stimuli by minimizing spurious, sensory-independent activity in auditory pathways.

### Hair cell activity is synchronized by ATP

Each ATP release event in Kölliker's organ is likely to affect multiple IHCs. We examined the relationship between the location and magnitude of these events and the amplitude of coincident spontaneous currents in IHCs (Fig. 6a, b). Spontaneous optical events that occurred near IHCs were associated with larger, faster rising currents (ratio of rise time to amplitude:  $61 \pm 3$  ms pA<sup>-1</sup>,  $n = 175$  events in 8 cells) than events occurring further away ( $89 \pm 6$  ms pA<sup>-1</sup>,  $n = 78$ ,  $P < 0.0001$ ). If IHCs exhibit similar sensitivity to ATP, neighbouring IHCs should experience similar activity. To determine whether the focal release of ATP is capable of synchronizing the output of neighbouring IHCs, we recorded spontaneous currents from pairs of IHCs located near (mean separation,  $d = 68 \pm 6 \mu\text{m}$ ) or far ( $d = 277 \pm 23 \mu\text{m}$ ) from one another. Whereas  $91 \pm 1\%$  of spontaneous inward currents occurred synchronously in neighbouring hair cells ( $n = 6$ ), only  $22 \pm 4\%$  ( $n = 6$ ,  $P < 0.0001$ ) were correlated in IHCs separated by a distance corresponding to a frequency change



**Figure 4 | Supporting-cell-derived ATP triggers bursts of action potentials in auditory nerve fibres.** **a**, Voltage-clamp recording from an afferent dendrite showing a burst of EPSCs elicited by ATP. Right: EPSC detail. **b**, Current-clamp recording from an afferent dendrite showing block of ATP-induced excitatory postsynaptic potentials by NBQX. **c**, Simultaneous recording of spontaneous EPSCs from an afferent dendrite (black trace) and spontaneous optical changes (blue trace) imaged in the surrounding region. Inset: EPSC detail from region marked by asterisk. **d**, Loose-patch recordings from two SGNs showing activity in response to ATP-dependent stimulation of the presynaptic IHC in control (black traces) and in the presence of nifedipine or NBQX (grey traces). The slow positive deflection of the field (source wave) results from activation of ATP receptors in supporting cells. **e**, Simultaneous SGN loose-patch recording (black trace) and spontaneous optical changes (blue trace) imaged in the region around the presynaptic IHC. Inset, waveform of action potential. **f**, Raster plot of spontaneous action potentials from eight SGNs, in the absence (cells 1–4) or presence (cells 5–8) of suramin (100–250  $\mu\text{M}$ ). Application of ACSF containing 30 mM KCl confirmed IHC responsiveness.



**Figure 5 | Intrinsic purinergic signalling ceases after the onset of hearing.** **a–c**, Spontaneous activity recorded from P7–10 (left) and P13–19 (right) cochleae. Spontaneous extracellular potentials (sEPs, **a**) and spontaneous optical changes (sOCs, **c**) were recorded in Kölliker's organ (left) or the inner sulcus (right). sICs, spontaneous inward currents. **d–f**, Plots of the frequency and amplitude of spontaneous extracellular potentials (**d**), spontaneous inward currents in IHCs (**e**) and spontaneous optical changes (**f**) before (age P7–10, pre) and after (age P13–19, post) hearing onset. The dashed lines indicate detection thresholds. Asterisk,  $P < 0.0001$ . Data in **d–f** represent mean  $\pm$  s.e.m. The number of experiments is indicated in parentheses.

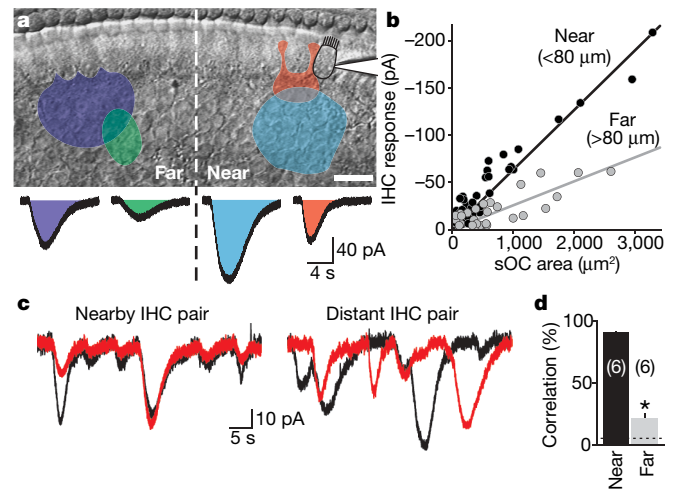
of one octave ( $\sim 250 \mu\text{m}$ )<sup>27</sup> (Fig. 6c, d). These results indicate that purinergic signalling in the developing organ of Corti is capable of synchronizing transmitter release from IHCs that will encode similar frequencies. If  $\text{Ca}^{2+}$  spikes are required to trigger glutamate release from these immature IHCs<sup>28,29</sup>, the number of auditory nerve fibres activated by each ATP-release event would be further restricted.

## Discussion

Spontaneous activity in auditory nerve fibres before the onset of hearing is essential for the survival of target neurons in the cochlear nucleus, accurate wiring of auditory pathways, and the refinement of tonotopic maps in auditory nuclei<sup>30–35</sup>. Our findings indicate that supporting cells within Kölliker's organ initiate bursts of electrical activity in SGNs before the onset of hearing through ATP-dependent excitation of hair cells. These results suggest that peripheral, non-sensory cells are essential for the maturation of auditory pathways in the brain. This activity may combine with intrinsically generated  $\text{Ca}^{2+}$  spikes<sup>29,36,37</sup> to produce distinct patterns of activity before the onset of hearing.

Synchronous activity among groups of hair cells along the length of the cochlea could help establish and maintain tonotopic segregation of neuronal projections in auditory pathways through hebbian-like plasticity<sup>31,38,39</sup>. Although correlated activity of neighbouring IHCs would be expected to reinforce connections to discrete targets in the central nervous system, by itself this activity would not provide information about the tonotopic position of active hair cells along the cochlea. It is likely that genetically encoded guidance cues are responsible for initial targeting and segregation of inputs in auditory nuclei, as has been described in the visual system<sup>40</sup>, whereas the periodic activity described here could help further refine and maintain synaptic connections<sup>34,41</sup>.

Although spontaneous release of ATP ceases after hearing onset, hair cells and supporting cells in the mature cochlea continue to express P2 receptors<sup>24,42</sup>, and cochlear injury triggers the release of



**Figure 6 | Local release of ATP synchronizes the activity of neighbouring IHCs.** **a**, Top: location and furthest extent of four spontaneous optical changes (pseudocolour overlay) recorded from one P9 organ of Corti. Bottom: coincident spontaneous inward currents recorded from one IHC (highlighted in top image). Scale:  $20 \mu\text{m}$ . **b**, Plot of peak amplitude of spontaneous inward currents from the IHC shown in **a**, versus maximum area of coincident spontaneous optical changes. Responses are grouped according to location relative to the IHC. Lines are regression fits. **c**, Spontaneous inward currents recorded simultaneously from two IHCs located near (left) or far from (right) one another. **d**, Plot showing percentage correlation of spontaneous inward currents recorded from IHCs located near (black) or far (grey) from one another. The dashed line indicates correlation predicted by chance alone. Asterisk,  $P < 0.0001$ . Data in **d** represent mean  $\pm$  s.e.m. The number of experiments is indicated in parentheses.

ATP and induces  $\text{Ca}^{2+}$  waves in supporting cells<sup>43</sup>. If spontaneous release of ATP were re-established under these conditions, sensory-independent activity may re-emerge in the auditory nerve, leading to conditions such as peripheral tinnitus. In addition, ATP is a known trophic factor, mitogen and potent neuromodulator<sup>44,45</sup>, raising the possibility that cells within Kölliker's organ participate in other vital aspects of cochlear development. A greater understanding of the mechanisms that regulate ATP release from cochlear supporting cells may help reveal new roles for this activity in the development and dysfunction of the auditory system.

## METHODS SUMMARY

Apical cochlear coils were removed from postnatal day (P) 7 to P19 Sprague–Dawley rats and used within 3 h of the dissection. Whole-cell, voltage- and current-clamp recordings from inner supporting cells, IHCs and auditory nerve afferents were performed under visual control, as described previously<sup>3,46</sup>. Spontaneous optical changes were imaged with DIC optics using a charge-coupled device (CCD) camera, and acquired at a rate of  $1 \text{ frame s}^{-1}$ . Transmittance changes were visualized by subtracting images 5 s apart, and quantified by applying a thresholding function to highlight changing pixels. Frequency measurements of spontaneous optical changes were normalized for a  $10^4 \mu\text{m}^2$  imaging area. Extracellular recordings of action potentials from SGNs and  $\text{Ca}^{2+}$  imaging were performed in cochlear organotypic cultures maintained for 1 to 7 days *in vitro*. Spontaneous and evoked changes in intracellular  $\text{Ca}^{2+}$  were monitored in cultures loaded with fura-2 or fluo-4 AM dyes, as described<sup>43</sup>. Statistical comparisons were made using paired or independent, two-tailed Student's *t*-tests assuming unequal variance, and significance was concluded when  $P < 0.05$ . Data are reported as mean  $\pm$  standard error of the mean (s.e.m.). Experiments were carried out in artificial cerebrospinal fluid (ACSF) at ambient temperature.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 1 June; accepted 10 September 2007.

- Geal-Dor, M., Freeman, S., Li, G. & Sohmer, H. Development of hearing in neonatal rats: air and bone conducted ABR thresholds. *Hear. Res.* **69**, 236–242 (1993).



2. Puel, J. L. & Uziel, A. Correlative development of cochlear action potential sensitivity, latency, and frequency selectivity. *Brain Res.* **465**, 179–188 (1987).
3. Glowatzki, E. & Fuchs, P. A. Transmitter release at the hair cell ribbon synapse. *Nature Neurosci.* **5**, 147–154 (2002).
4. Gummer, A. W. & Mark, R. F. Patterned neural activity in brain stem auditory areas of a prehearing mammal, the tammar wallaby (*Macropus eugenii*). *Neuroreport* **5**, 685–688 (1994).
5. Jones, T. A., Jones, S. M. & Pagggett, K. C. Primordial rhythmic bursting in embryonic cochlear ganglion cells. *J. Neurosci.* **21**, 8129–8135 (2001).
6. Jones, T. A., Leake, P. A., Snyder, R. L., Stakhovskaya, O. & Bonham, B. H. Spontaneous discharge patterns in cochlear spiral ganglion cells prior to the onset of hearing in cats. *J. Neurophysiol.* doi:10.1152/jn.00472.2007 (8 August 2007).
7. Walsh, E. J. & McGee, J. Postnatal development of auditory nerve and cochlear nucleus neuronal responses in kittens. *Hear. Res.* **28**, 97–116 (1987).
8. Lippe, W. R. Rhythmic spontaneous activity in the developing avian auditory system. *J. Neurosci.* **14**, 1486–1495 (1994).
9. Retzius, G. *Das Gehörorgan der Wirbelthiere. II. Das Gehörorgan der Reptilien, der Vögel und Säugethiere* (Samson & Wallin, Stockholm, 1884).
10. Hinojosa, R. A note on development of Corti's organ. *Acta Otolaryngol. (Stockh.)* **84**, 238–251 (1977).
11. Wada, T. Anatomical and physiological studies on the growth of the inner ear of the albino rat. *Am. Anat. Mem.* **10**, 1–174 (1923).
12. Eybalin, M. Neurotransmitters and neuromodulators of the mammalian cochlea. *Physiol. Rev.* **73**, 309–373 (1993).
13. Glowatzki, E. & Fuchs, P. A. Cholinergic synaptic inhibition of inner hair cells in the neonatal mammalian cochlea. *Science* **288**, 2366–2368 (2000).
14. Fuchs, P. A., Evans, M. G. & Murrow, B. W. Calcium currents in hair cells isolated from the cochlea of the chick. *J. Physiol. (Lond.)* **429**, 553–568 (1990).
15. Burnstock, G. Physiology and pathophysiology of purinergic neurotransmission. *Physiol. Rev.* **87**, 659–797 (2007).
16. Housley, G. D. Physiological effects of extracellular nucleotides in the inner ear. *Clin. Exp. Pharmacol. Physiol.* **27**, 575–580 (2000).
17. North, R. A. Molecular physiology of P2X receptors. *Physiol. Rev.* **82**, 1013–1067 (2002).
18. Bennett, M. V., Contreras, J. E., Bukauskas, F. F. & Saez, J. C. New roles for astrocytes: gap junction hemichannels have something to communicate. *Trends Neurosci.* **26**, 610–617 (2003).
19. Forge, A. *et al.* Gap junctions in the inner ear: comparison of distribution patterns in different vertebrates and assessment of connexin composition in mammals. *J. Comp. Neurol.* **467**, 207–231 (2003).
20. Zhao, H. B. Connexin26 is responsible for anionic molecule permeability in the cochlea for intercellular signalling and metabolic communications. *Eur. J. Neurosci.* **21**, 1859–1868 (2005).
21. Charles, A. & Giaume, C. *Intercellular Calcium Waves in Astrocytes: Underlying Mechanisms and Functional Significance* (eds Volterra, A., Magistretti, P. J. & Haydon, P. G.) (Oxford Univ. Press, Oxford, 2002).
22. Piazza, V., Ciubotaru, C. D., Gale, J. E. & Mammano, F. Purinergic signalling and intercellular  $\text{Ca}^{2+}$  wave propagation in the organ of Corti. *Cell Calcium* **41**, 77–86 (2007).
23. Salih, S. G., Jagger, D. J. & Housley, G. D. ATP-gated currents in rat primary auditory neurones *in situ* arise from a heteromultimeric P2X receptor subunit assembly. *Neuropharmacology* **42**, 386–395 (2002).
24. Sugawara, M., Erostequi, C., Blanchet, C. & Dulon, D. ATP activates non-selective cation channels and calcium release in inner hair cells of the guinea-pig cochlea. *J. Physiol. (Lond.)* **491**, 707–718 (1996).
25. Jones, T. A. & Jones, S. M. Spontaneous activity in the statoacoustic ganglion of the chicken embryo. *J. Neurophysiol.* **83**, 1452–1468 (2000).
26. Rübsamen, R. & Lippe, W. R. in *Development of the Auditory System* (eds Rubel, E. W., Popper, A. N. & Fay, R. R.) 193–270 (Springer, New York, 1998).
27. Müller, M. Frequency representation in the rat cochlea. *Hear. Res.* **51**, 247–254 (1991).
28. Beutner, D. & Moser, T. The presynaptic function of mouse cochlear inner hair cells during development of hearing. *J. Neurosci.* **21**, 4593–4599 (2001).
29. Marcotti, W., Johnson, S. L., Rüsch, A. & Kros, C. J. Sodium and calcium currents shape action potentials in immature mouse inner hair cells. *J. Physiol. (Lond.)* **552**, 743–761 (2003).
30. Friauf, E. & Lohmann, C. Development of auditory brainstem circuitry. Activity-dependent and activity-independent processes. *Cell Tissue Res.* **297**, 187–195 (1999).
31. Kandler, K. Activity-dependent organization of inhibitory circuits: lessons from the auditory system. *Curr. Opin. Neurobiol.* **14**, 96–104 (2004).
32. Rubel, E. W. & Fritzsch, B. Auditory system development: primary auditory neurons and their targets. *Annu. Rev. Neurosci.* **25**, 51–101 (2002).
33. Leao, R. N. *et al.* Topographic organization in the auditory brainstem of juvenile mice is disrupted in congenital deafness. *J. Physiol. (Lond.)* **571**, 563–578 (2006).
34. Leake, P. A., Hradek, G. T., Chair, L. & Snyder, R. L. Neonatal deafness results in degraded topographic specificity of auditory nerve projections to the cochlear nucleus in cats. *J. Comp. Neurol.* **497**, 13–31 (2006).
35. Gabriele, M. L., Brunso-Bechtold, J. K. & Henkel, C. K. Plasticity in the development of afferent patterns in the inferior colliculus of the rat after unilateral cochlear ablation. *J. Neurosci.* **20**, 6939–6949 (2000).
36. Kros, C. J., Rüppersberg, J. P. & Rüsch, A. Expression of a potassium current in inner hair cells during development of hearing in mice. *Nature* **394**, 281–284 (1998).
37. Marcotti, W., Johnson, S. L., Holley, M. C. & Kros, C. J. Developmental changes in the expression of potassium currents of embryonic, neonatal and mature mouse inner hair cells. *J. Physiol. (Lond.)* **548**, 383–400 (2003).
38. Katz, L. C. & Shatz, C. J. Synaptic activity and the construction of cortical circuits. *Science* **274**, 1133–1138 (1996).
39. Kotak, V. C. & Sanes, D. H. Synaptically evoked prolonged depolarizations in the developing auditory system. *J. Neurophysiol.* **74**, 1611–1620 (1995).
40. Huberman, A. D. Mechanisms of eye-specific visual circuit development. *Curr. Opin. Neurobiol.* **17**, 73–80 (2007).
41. Erazo-Fischer, E., Striessnig, J. & Taschenberger, H. The role of physiological afferent nerve activity during *in vivo* maturation of the calyx of Held synapse. *J. Neurosci.* **27**, 1725–1737 (2007).
42. Lagostena, L. & Mammano, F. Intracellular calcium dynamics and membrane conductance changes evoked by Deiters' cell purinoceptor activation in the organ of Corti. *Cell Calcium* **29**, 191–198 (2001).
43. Gale, J. E., Piazza, V., Ciubotaru, C. D. & Mammano, F. A mechanism for sensing noise damage in the inner ear. *Curr. Biol.* **14**, 526–529 (2004).
44. Fields, R. D. & Burnstock, G. Purinergic signalling in neuron–glia interactions. *Nature Rev. Neurosci.* **7**, 423–436 (2006).
45. Nedergaard, M., Ransom, B. & Goldman, S. A. New roles for astrocytes: redefining the functional architecture of the brain. *Trends Neurosci.* **26**, 523–530 (2003).
46. Glowatzki, E. *et al.* The glutamate-aspartate transporter GLAST mediates glutamate uptake at inner hair cell afferent synapses in the mammalian cochlea. *J. Neurosci.* **26**, 7659–7664 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J.-H. Kong for help with preliminary experiments, and P. Fuchs, J. Howell and M. Lahne for discussions. This work was supported by a Royal Society University Research Fellowship (to J.E.G.), National Institutes of Health Grants (to E.G. and D.E.B.) and the Deafness Research Foundation (to D.E.B.).

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to [dbergles@jhmi.edu](mailto:dbergles@jhmi.edu).

## METHODS

**Tissue preparation.** All experimental protocols were approved by the Johns Hopkins University Animal Care and Use Committee. Apical turns of postnatal day (P) 7 to 19 (day of birth is P0) Sprague–Dawley rat cochleae were isolated in ice-cold artificial cerebral spinal fluid (ACSF) containing (in mM): NaCl (119), KCl (2.5), CaCl<sub>2</sub> (2.5), MgCl<sub>2</sub> (1.3), NaH<sub>2</sub>PO<sub>4</sub> (1), NaHCO<sub>3</sub> (26.2) and glucose (11), saturated with 95% O<sub>2</sub>/5% CO<sub>2</sub>. Other than the stria vascularis and tectorial membrane, which were gently peeled off, the cellular organization of the organ of Corti was left intact. Acutely isolated cochlear turns were transferred to a Plexiglas chamber mounted on an upright microscope (Zeiss Axioskop FS2) and imaged using infrared light (~770 nm) and DIC optics using a charge-coupled device (CCD) camera (Sony XC-73). Organs of Corti were continually superfused with ACSF at 22–24 °C, and used within 3 h of the dissection. For afferent dendrite recordings, the bath solution contained (in mM): NaCl (144), KCl (5.8), CaCl<sub>2</sub> (1.3), MgCl<sub>2</sub> (0.9), NaH<sub>2</sub>PO<sub>4</sub> (0.7), HEPES (10) and glucose (5.6), pH 7.4. For organotypic cultures, P1–7 cochlear turns were plated on Cell-Tak-coated (BD Biosciences) coverslips and maintained for 1 to 7 days *in vitro* in F12/DMEM (Invitrogen) supplemented with 1% fetal bovine serum. Ca<sup>2+</sup> imaging was performed as described previously<sup>43</sup>.

**Electrophysiology.** Whole-cell, voltage- and current-clamp recordings from supporting cells, IHCs and auditory nerve dendrites were performed under visual control as described previously<sup>3,46</sup>. Electrodes were advanced through the tissue under minimal positive pressure to limit cellular disruption. Intracellular solutions were composed of the following (in mM): KCH<sub>3</sub>SO<sub>3</sub> (134), HEPES (20), EGTA (10), MgCl<sub>2</sub> (1), pH 7.3. For IHC recordings, the electrode solution was supplemented with Na<sub>2</sub>ATP (2 mM) and NaGTP (0.2 mM) to allow long and stable recordings. In the presence of 2.5 mM KCl, IHCs in P7–10 organs of Corti had a resting membrane potential ( $V_m$ ) of approximately –75 mV and did not fire tonically. For afferent recordings, the internal solution contained (in mM): KCl (150), MgCl<sub>2</sub> (3.5), CaCl<sub>2</sub> (0.1), EGTA (5), HEPES (5), pH 7.2. Nucleotides were omitted from most internal solutions to minimize P2 receptor desensitization (see Supplementary Fig. 3). Extracellular action potentials were recorded from SGNs by making low-resistance seals (~15 M $\Omega$ ) with their somata. Pipette resistances were 1–2 M $\Omega$  for field- and loose-patch recordings, 2–4 M $\Omega$  for whole-cell recordings, and 10–15 M $\Omega$  for afferent dendrite recordings. During voltage-clamp recordings, the membrane potentials of IHCs, afferent dendrites, and supporting cells were held at –80 mV, –84 mV and –90 mV, respectively, unless otherwise indicated. Errors due to the voltage drop across the series resistance were left uncompensated. Membrane voltage and currents were recorded with pClamp9 software using a Multiclamp 700A amplifier, low-pass filtered at 0.1–10 kHz, and digitized at 10–50 kHz with a Digidata 1322A analogue to digital converter (Molecular Devices). Data were analysed off-line using Clampfit (Molecular Devices) and Origin (Microcal Software) software. Spontaneous extracellular potentials and supporting cell

spontaneous inward currents with amplitudes above baseline noise (0.03 mV and 8 pA, respectively) were measured and deemed correlated when they occurred within  $\pm 250$  ms. The mean field potential amplitude was calculated from absolute values. IHC currents were considered coincident when their peaks occurred within  $\pm 500$  ms of one another. EPSC and action potential bursts were defined as groups of a minimum of four events occurring at a frequency greater than 2.4 Hz (mean frequency:  $8.14 \pm 1.11$  Hz,  $n = 46$ ) that differed significantly from the mean EPSC frequency between bursts ( $0.17 \pm 0.03$  Hz,  $P < 0.0001$ ). For pharmacological studies, a stable baseline was recorded for 20 min, then one of the following drugs (all from Sigma, unless specified otherwise) was superfused for 10 min: PPADS (50  $\mu$ M; Tocris), suramin (100–250  $\mu$ M, unless indicated otherwise; Tocris), octanol (1 mM), carbenoxolone (500  $\mu$ M), strychnine (1  $\mu$ M); D-tubocurarine (10  $\mu$ M), atropine (3  $\mu$ M), TTX (1  $\mu$ M; Alamone), nifedipine (50  $\mu$ M; Tocris), NBQX (10  $\mu$ M; Tocris), R,S-CPP (10  $\mu$ M; Tocris), DPCPX (5  $\mu$ M; Tocris), TNP-ATP (1  $\mu$ M; Tocris), brilliant blue G (10  $\mu$ M), CdCl<sub>2</sub> (30  $\mu$ M), followed by a 20 to 30 min recovery period. Apyrase (50 U ml<sup>–1</sup>; Sigma) was applied for 30 min under static bath in HEPES-buffered ACSF, pH 7.4. In each case, the mean absolute amplitude and frequency were calculated for 200-s bins, normalized to baseline averages, and for statistical analysis compared to values obtained at corresponding times in control preparations continuously bathed in ACSF. Only one drug was applied per cochlear turn. For focal application of nucleotides, the agonists were dissolved in HEPES-buffered ACSF to a concentration of 100  $\mu$ M, loaded into a patch pipette, and expelled from the tip using a Picospritzer (50–100 ms, 5 p.s.i.). Statistical comparison of means was made using paired or independent, two-tailed Student's *t*-tests assuming unequal variance, and significance was concluded when  $P < 0.05$ . Data are presented as mean  $\pm$  s.e.m.

**Intrinsic optical imaging.** Cochlear turns were imaged with a  $\times 40$  water immersion objective coupled to an additional  $\times 0.5$ – $2.0$  adjustable zoom lens (Zeiss) to allow low-magnification imaging of 220- $\mu$ m-long cochlear segments after whole-cell recordings were established at high-magnification. IR-DIC images were acquired at 1 frame per second using an LG-3 frame grabber card (Scion Corporation) and Scion Image software. Difference movies were created by subtracting frames captured at times  $t_n$  and  $t_{n+5}$  seconds from each other using ImageJ software, to provide an index of transmittance change ( $\Delta T$ ) over time: stable regions are depicted in grey, whereas regions undergoing spontaneous optical changes appear black and white. To quantify intrinsic optical changes, a thresholding function was applied to highlight and count pixels that changed (in arbitrary units, AU). Identical contrast and threshold values were used to quantify all movies in a given experiment. Spontaneous optical change frequency measurements were normalized for a  $10^4 \mu\text{m}^2$  imaging area. Plots of transmittance change over time were precisely aligned to electrophysiological traces by taking into consideration the 5-s lag artificially created through the image subtraction process.



## ARTICLES

# Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system

Jean Livet<sup>1</sup>, Tamily A. Weissman<sup>1</sup>, Hyuno Kang<sup>1</sup>, Ryan W. Draft<sup>1</sup>, Ju Lu<sup>1</sup>, Robyn A. Bennis<sup>1</sup>, Joshua R. Sanes<sup>1</sup> & Jeff W. Lichtman<sup>1</sup>

**Detailed analysis of neuronal network architecture requires the development of new methods. Here we present strategies to visualize synaptic circuits by genetically labelling neurons with multiple, distinct colours. In *Brainbow* transgenes, *Cre/lox* recombination is used to create a stochastic choice of expression between three or more fluorescent proteins (XFPs). Integration of tandem *Brainbow* copies in transgenic mice yielded combinatorial XFP expression, and thus many colours, thereby providing a way to distinguish adjacent neurons and visualize other cellular interactions. As a demonstration, we reconstructed hundreds of neighbouring axons and multiple synaptic contacts in one small volume of a cerebellar lobe exhibiting approximately 90 colours. The expression in some lines also allowed us to map glial territories and follow glial cells and neurons over time *in vivo*. The ability of the *Brainbow* system to label uniquely many individual cells within a population may facilitate the analysis of neuronal circuitry on a large scale.**

Cajal revolutionized neurobiology when he used Golgi's silver stain to label small numbers of neurons in their entirety, thereby identifying the cellular elements of neural circuits. The small number of labelled cells, however, was also a limitation because quantitative information such as divergence and convergence at synaptic relays was inaccessible. Efforts are presently underway to produce connectivity maps in which multiple, or even all, neuronal connections are rendered<sup>1–3</sup>. Building such 'connectomic' maps would be more straightforward with the equivalent of a multicolour Golgi stain that would allow many neurons within a single sample to be individually identified by virtue of a large number of cell-specific labels. In the same way that a television monitor encodes colour space by mixing three primary channels—red, green and blue—the combination of three (or more) coloured dyes can generate many different hues<sup>4</sup>. Multiple spectral variants of fluorescent proteins now exist<sup>5</sup> and are ideal labels for this purpose.

Combining two XFPs in one animal and stably expressing them in a mosaic manner has been achieved by various strategies, including the generation of chimaeric mice mixing stem cells of two colours<sup>6</sup>; the selection and crossing of mouse lines showing variegated expression of yellow fluorescent protein (YFP) and cyan fluorescent protein (CFP)<sup>7,8</sup>; or a genetic scheme reconstituting split XFP genes by crossing over<sup>9</sup>. The first two approaches are, however, cumbersome, and all three generate only a limited palette of colours.

To circumvent these limitations, we exploited the advantages of the widely used *Cre/lox* recombination system<sup>10</sup>, which can switch on gene expression by DNA excision, inversion, or interchromosomal recombination<sup>9–11</sup>. We designed two genetic strategies, called *Brainbow*, for stochastic expression of multiple fluorescent proteins from a single transgene. *Brainbow-1* uses *Cre*-mediated excision between pairs of incompatible *lox* sites, alternated to create mutually exclusive recombination events. In *Brainbow-2*, *Cre* inverts DNA segments delimited by *loxP* sites in opposite orientation, positioned in tandem to generate several recombination outcomes. The differential

expression of multiple copies of these constructs generates XFP mixtures, allowing the labelling of individual neurons and glia with as many as 90 distinguishable colours.

## Strategies for mosaic expression of multiple genes

**Stochastic recombination using incompatible *lox* sites.** *Lox* variants have been developed with mutations that prevent them from recombining with the canonical *loxP* site, while allowing them to recombine specifically with identical *lox* sequences<sup>10,12</sup>. We reasoned that by alternating such variant *lox* sites with canonical *loxP* in the same construct, we could force *Cre* to choose between two mutually exclusive excision events (Fig. 1a). In this configuration, which we call *Brainbow-1*, two recombination events are initially possible, but only one can occur. Excision between either pair of identical *lox* sites removes one of the other pair, thereby preventing further recombination.

***Brainbow-1.0.*** We first verified that the *lox* variant *lox2272* (ref. 12 and Supplementary Table 1), which is incompatible with *loxP* in prokaryotic systems, would be comparably selective in mammalian cells (Supplementary Fig. 1a, b). We then built a construct, *Brainbow-1.0*, in which the arrangement of *loxP* and *lox2272* allows two possible recombination outcomes, switching XFP expression from default red to either yellow or cyan (Fig. 1a). As expected, HEK 293 cells stably transfected with the construct expressed only red fluorescent protein (RFP) from the cDNA immediately adjacent to the promoter. On *Cre* action, many cells lost RFP expression and switched on YFP or CFP, creating a three-colour cellular array (Fig. 1b). Yellow and cyan cells were present in comparable numbers ( $50 \pm 15$  YFP<sup>+</sup> cells and  $52 \pm 12$  CFP<sup>+</sup> cells per field), suggesting that neither recombination pathway is strongly favoured over the other. Co-expression of YFP and CFP was not observed in the cell line shown in Fig. 1b, but was seen in other clones, presumably because of the presence of multiple copies of the transgene (see below).

***Brainbow-1.1.*** To permit expression of a fourth XFP, we designed an additional *lox* variant, the efficiency of which was equivalent to *loxP*

<sup>1</sup>Department of Molecular and Cellular Biology and Center for Brain Science, Harvard University, Cambridge, Massachusetts 02138, USA.

and *lox2272* but was incompatible with both of them (Supplementary Fig. 1c, d). This new variant, *loxN*, has substitutions in the *lox* spacer that differ from both *loxP* and *lox2272* (Supplementary Table 1). Using *loxN*, we built a *Brainbow-1.1* construct bearing orange (OFP), red, yellow and cyan fluorescent protein genes, where Cre is offered a choice between three recombination events (Fig. 1c). In the absence of recombination, HEK cell lines expressing this construct showed OFP fluorescence, whereas Cre triggered expression of the other XFPs (Fig. 1d).

**Stochastic recombination using Cre-mediated inversion.** Cre can invert DNA segments flanked by *loxP* sites facing each other. Inversion can continue for as long as the recombinase is present, but restricting Cre activity to a limited time stabilizes the transgene in one or the other orientation independently in each cell<sup>11</sup>. In *Brainbow-2*, we use inversion to obtain multiple XFP expression.

**Brainbow-2.0.** We first built a Cre-invertible construct containing RFP and CFP in head-to-head orientation (Fig. 2a). In HEK cells, this construct, *Brainbow-2.0*, drove expression of the sense-oriented RFP. Transient transfection with Cre activated expression of the antisense-oriented CFP in some cells, indicating that the inversion strategy was effective (Fig. 2b).

**Brainbow-2.1.** On the basis of the success of *Brainbow-2.0*, we designed *Brainbow-2.1*, in which two invertible DNA segments are positioned in tandem to generate a larger number of recombination outcomes (Fig. 2c). Three different inversions can occur in *Brainbow-2.1* (see i, ii and iii in Fig. 2c). In addition, either of two excision events (iv, v) may reduce the construct to one of two single invertible DNA segments, which can continue to invert as long as Cre is present. These different recombination possibilities can be used to express four genes. HEK cells stably transfected with a *Brainbow-2.1* construct containing green, yellow, red and cyan fluorescent protein

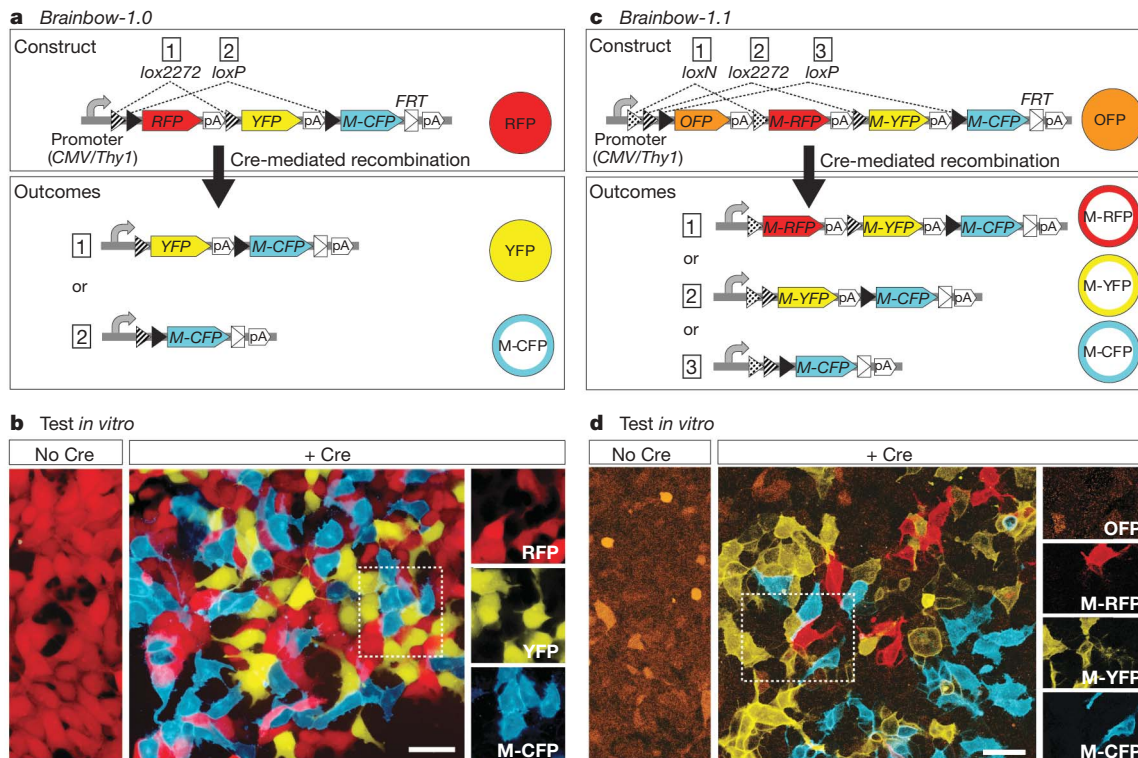
showed mosaic expression of all XFPs after transient transfection with Cre (Fig. 2d).

Thus, stochastic expression of up to four XFPs can be obtained from a single *Brainbow-1* or *Brainbow-2* transgene in which the Cre recombinase is offered a choice of excision and/or inversion between multiple pairs of *lox* sites.

### Brainbow in vivo

**Thy1-Brainbow transgenic mice.** To generate lines of mice expressing *Brainbow* in neurons, the *Brainbow-1* and *Brainbow-2* constructs were placed under the control of regulatory elements from the *Thy1* gene, which are known to drive expression at high levels in a variety of neuron types<sup>7,13</sup>. Nineteen *Thy1-Brainbow* lines showing transgene expression were obtained: twelve with the *Brainbow-1.0* construct, one with *Brainbow-1.1*, two with *Brainbow-2.0* and four with *Brainbow-2.1* (Supplementary Table 2). The labelled cell types were similar to those seen previously in *Thy1* transgenic animals<sup>7,13</sup>, although expression in glial cells was also observed here in four lines (see below).

To induce recombination in the *Thy1-Brainbow* mouse lines, we crossed them with *CAGGS-CreERT2* animals<sup>14</sup>. In these mice, *CAGGS* drives broad expression of an oestrogen receptor–Cre fusion specifically activated by injection of the ligand tamoxifen<sup>15</sup>. We analysed double heterozygous adult animals (*CAGGS-CreERT2*<sup>+/−</sup>; *Thy1-Brainbow*<sup>+/−</sup>) that had been injected with a single dose of tamoxifen at postnatal day 0–3 (P0–P3) to activate recombination in cells co-expressing *Brainbow* and CreER. Recombination leading to new colours was observed in all 19 *Brainbow* lines (Fig. 3a–d). In most lines, the number and location of expressing cells appeared unchanged after recombination. However, in some lines the population of cells expressing XFPs was larger in the recombined animals than in the non-recombined controls (Supplementary Table 2). This



**Figure 1 | Brainbow-1: stochastic recombination using incompatible *lox* variants.** **a**, In *Brainbow-1.0*, incompatible sets of *lox* sites alternate: Cre chooses between excision events 1 or 2. Before Cre action, only the gene following the promoter is expressed (RFP). Recombination switches expression to either YFP (1) or M-CFP (2). **b**, HEK cells stably transfected with *CMV-Brainbow-1.0* express RFP. On transient transfection with Cre, these cells randomly switch to YFP or M-CFP expression. **c**, In *Brainbow-1.1*,

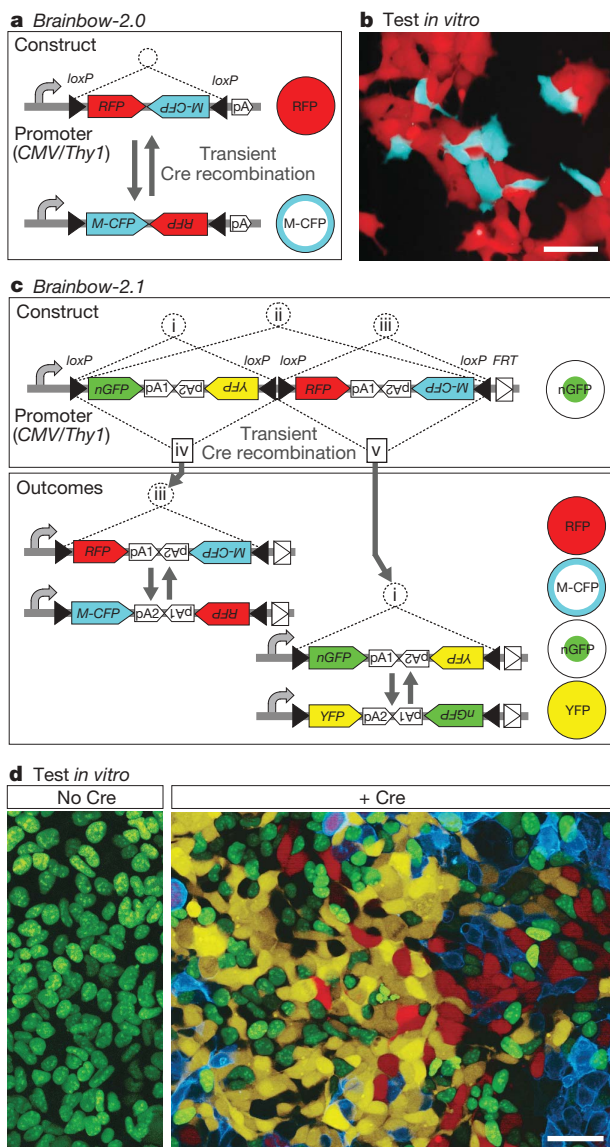
a third set of incompatible *lox* sites (*loxN*) is added, creating three recombination possibilities (1, 2 or 3), switching OFP expression to RFP, YFP or CFP expression. **d**, Cells stably transfected with *Brainbow-1.1* express OFP. Cre recombination leads to expression of M-RFP, M-YFP or M-CFP. pA, polyadenylation signal; M-XFP, membrane-tethered XFP. *FRT* site allows reduction of transgene arrays (Fig. 4d). Scale bar, 50  $\mu$ m.



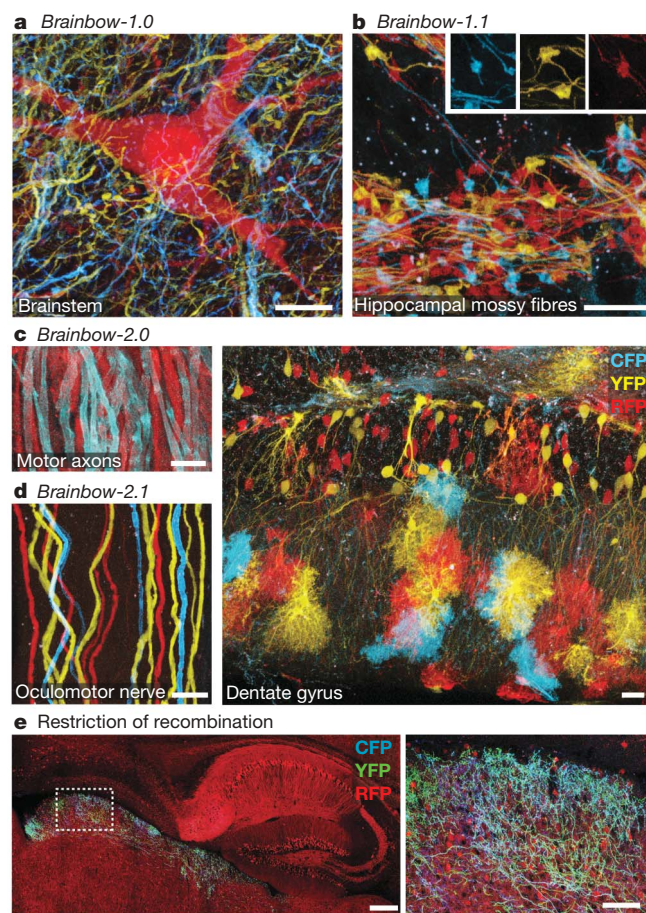
altered expression is consistent with the observation that shortening a transgene can increase its probability of expression<sup>16</sup>. These tests show that recombination of both *Brainbow-1* and *Brainbow-2* transgenes leads to mosaic expression of multiple fluorescent proteins *in vivo*. As was the case *in vitro*, no XFP expression other than the default was detected in the absence of Cre (Supplementary Fig. 3a). **Combinatorial XFP expression.** In some *Thy1-Brainbow* mouse lines, individual neurons expressed only one of the XFPs present in the transgene (Fig. 3d and Supplementary Table 2). More commonly however, transgenic *Brainbow* lines showed co-expression of multiple colours in individual cells (Figs 4 and 5, Supplementary Fig. 2 and Supplementary Table 2). Notably, a large number of different hues was observed. In a confocal stack through a peripheral motor nerve, for instance, we distinguished at least ten XFP combinations

in different axons (Fig. 4b). Neurons of the central nervous system also exhibited combinatorial XFP expression (Figs 4c and 5, and Supplementary Fig. 2).

How were these colour combinations generated? Polychromatic cells were not unexpected given that pronuclear injection typically leads to the tandem integration of multiple transgene copies<sup>17</sup>. If, for example, three copies of the *Brainbow-1.0* construct were present in mice and recombined independently, at least ten colour mixtures would be expected (Fig. 4a). Indeed, quantitative polymerase chain reaction (PCR) tests detected eight transgene copies in *Brainbow-1.0* line H and 16 in line G (data not shown). To investigate whether combinatorial XFP expression was caused by these multiple copies, we reduced their number<sup>16</sup>, making use of an *FRT* site that we had inserted into the *Brainbow* transgene (Fig. 1a). Applying Flp/*FRT*-mediated recombination (analogous to the Cre/*lox* system but using a different recombinase and target sequence<sup>10</sup>), we excised tandem repeats of the *Brainbow* construct by crossing lines G and H with Flp-expressing mice<sup>18</sup> (Fig. 4d). Two *Thy1-Brainbow-1.0*<sup>(Flp)</sup> lines were

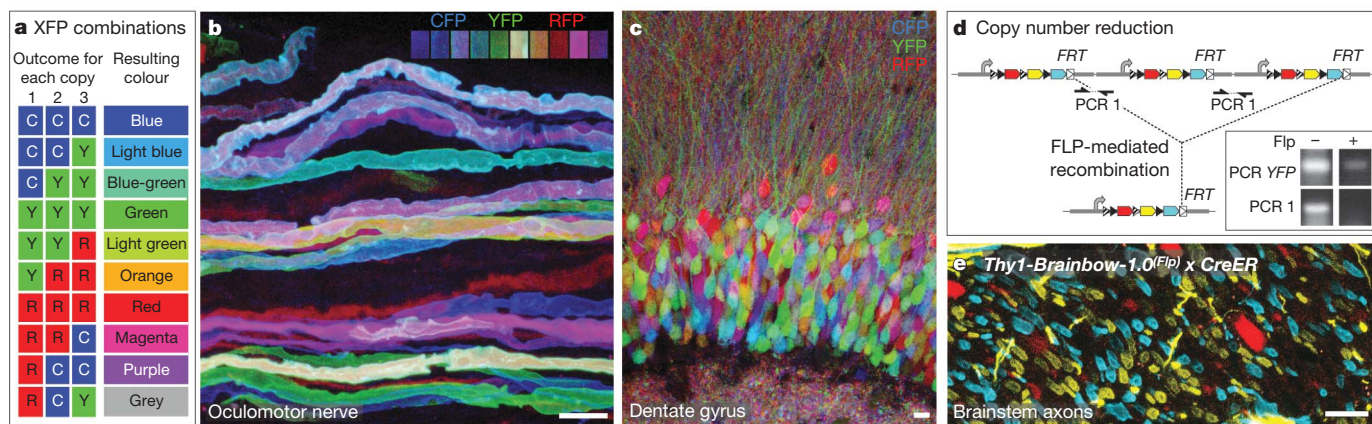


**Figure 2 | Brainbow-2: stochastic recombination using Cre-mediated inversion.** **a**, In *Brainbow-2.0*, Cre triggers inversion of a DNA segment flanked by *loxP* sites in opposite orientation. In 50% of cells, inversion should end in an antisense orientation and switch gene expression. **b**, HEK cells stably expressing CMV-*Brainbow-2.0* produce RFP, and stochastically switch to CFP expression when transfected with Cre. **c**, The *Brainbow-2.1* construct contains two tandem invertible DNA segments. Inversion (i-iii) and excision (iv, v) recombination events create four expression possibilities. **d**, Stable CMV-*Brainbow-2.1* transfectants express nuclear GFP (nGFP). Cre recombination triggers expression of YFP, RFP or M-CFP. pA1 and pA2, SV40 and bGH polyadenylation signals. Scale bars, 50  $\mu$ m.



**Figure 3 | XFP expression in Brainbow transgenic mice.** **a**, **b**, *Thy1-Brainbow-1.0* and *Thy1-Brainbow-1.1* transgenic mice were crossed with CreERT2-expressing animals. Tamoxifen injection led to mosaic XFP expression throughout the brain. **a**, Brainstem, line H; **b**, hippocampal mossy fibre axons and their terminals (see insets), line M. **c**, In *Thy1-Brainbow-2.0* mice, transient recombination with the CreERT2/tamoxifen system triggers expression of M-CFP (peripheral motor axons, line N). **d**, In *Thy1-Brainbow-2.1* mice, CreERT2-mediated recombination leads to expression of multiple XFPs. Left: oculomotor nerve, line R. Right: hippocampus (dentate gyrus), line Q (labelled neurons and astrocytes). **e**, Sagittal brain sections of *Thy1-Brainbow-1.0* mice line H crossed with the retina-specific *Chx10-Cre* driver. Recombination is almost completely restricted to retinal ganglion cells, as shown by label of their axons arborizing in the superior colliculus. The boxed area in the left panel is shown at greater magnification in the right panel. Scale bars: **a-c**, 10  $\mu$ m; **d**, 20  $\mu$ m; **e** (left), 250  $\mu$ m; **e** (right), 50  $\mu$ m.





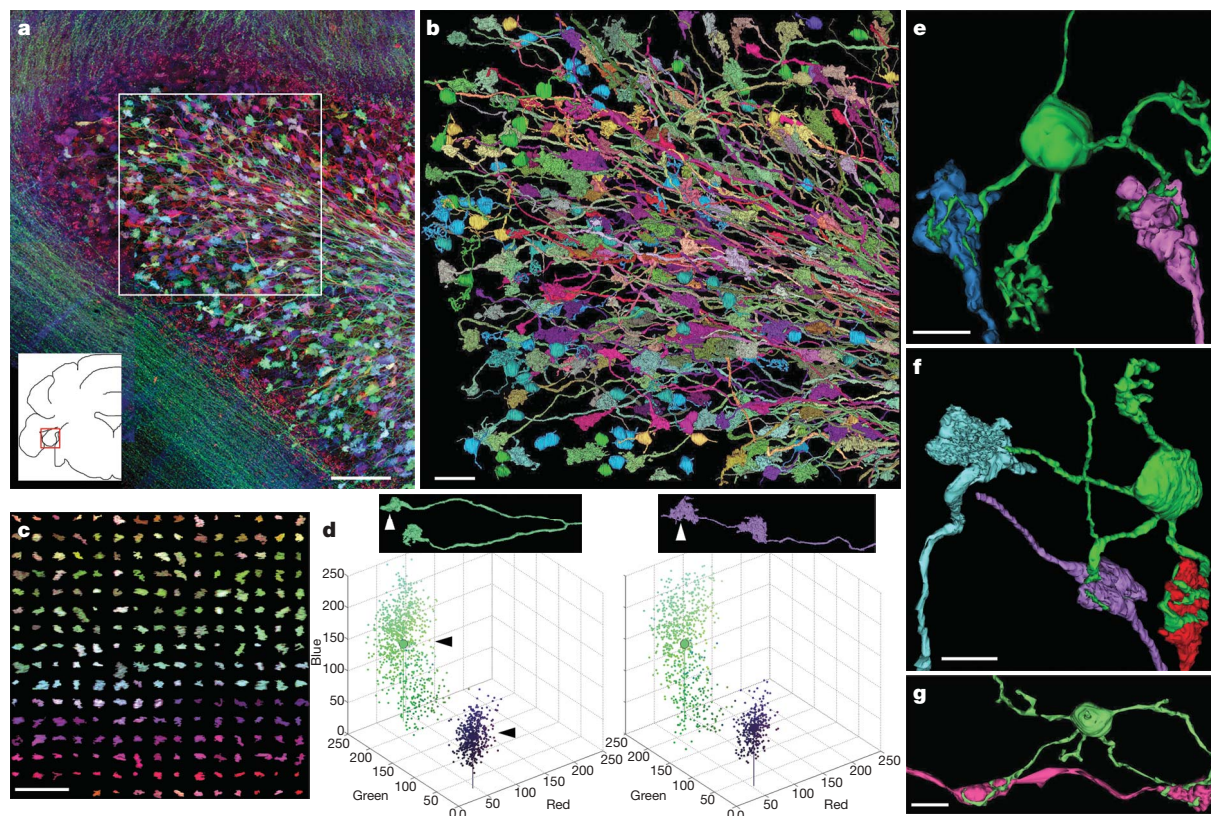
**Figure 4 | Combinatorial XFP expression results from tandem copy integration.** **a**, With a *Brainbow* construct expressing three XFPs, independent recombination of three transgene copies can, in principle, generate ten distinct colour combinations. **b**, Oculomotor axons of *Thy1-Brainbow-1.0* line H (recombination with CreERT2). Boxes show sample regions from different axons. **c**, Dentate gyrus of *Thy1-Brainbow-1.0* line L

(recombination with CreERT2). **d**, A single *FRT* site inserted in *Brainbow* constructs allows tandem transgene copy number reduction through Flp-mediated recombination. The PCR indicates the disappearance of transgene repeats in *Thy1-Brainbow-1.0* line H crossed with Flp-expressing mice (inset). **e**, A Flp-recombined line derived from line H expresses XFPs in a mutually exclusive manner. Scale bars, 10  $\mu$ m.

derived that showed mutually exclusive rather than combinatorial XFP expression in all animals tested ( $n = 5$  for each line, Fig. 4e).

Thus, combinatorial gene expression in *Brainbow* mice appears to be generated by multiple copies of the transgene recombining independently towards different outcomes. Partial inter-copy recombination might also contribute to the palette of colour observed.

**Recombination with specific Cre drivers.** In principle, stochastic expression from the excision-based *Brainbow-1* transgene could be activated by chronically active Cre rather than via tamoxifen-inducible CreER. To test this idea, we crossed *Thy1-Brainbow-1.0* mice showing neuronal expression to *Chx10* (also called *Vsx2*)-Cre mice, which drive recombination in the retina<sup>19</sup>. *Brainbow* expression



**Figure 5 | Cerebellar circuit tracing and colour analysis.** **a**, Cerebellar flocculus from line H. Inset shows coronal location. **b**, Three-dimensional digital reconstruction of region boxed in **a** (341 axons and 93 granule cells; 160  $\mu$ m<sup>2</sup>  $\times$  65  $\mu$ m). (See also Supplementary Movie 1.) **c**, Colour distribution of rosettes. Each rosette from the reconstructed region is aligned according to hue. **d**, Colour constancy along axon. Top: two mossy fibre axons (green, purple) each possess two presynaptic rosettes. Bottom: pixel distribution for each rosette (R, G and B intensities displayed on a scale

from 0 to 255). The left graph displays upper rosettes for each axon (arrowheads); right graph displays corresponding lower rosette. **e**, Reconstructed granule cell receives input from  $\geq 3$  different mossy fibres (blue, pink and at least 1 unlabelled). The granule cell axon projects upwards. **f**, Each granule cell dendrite is innervated by a different presynaptic neuron (three labelled, one unlabelled). **g**, Two granule cell dendrites are innervated by the same presynaptic mossy fibre. Scale bars: **a**, **c**, 50  $\mu$ m; **b**, 15  $\mu$ m; **e**–**g**, 5  $\mu$ m.



in the brain (YFP, CFP and combinations) was restricted to retinal projections (Fig. 3e). Additional Cre lines tested similarly drove *Brainbow* expression in territories where expression of *Thy1* and Cre intersected (Supplementary Fig. 3).

### Circuit tracing with Brainbow

To determine whether combinatorial XFP expression in *Brainbow* mice can be used as a tool for mapping neural circuits, we studied the inner granular layer (IGL) of the cerebellum. *Brainbow-1.0* line H labelled two interconnected components of IGL circuitry: subsets of mossy fibre axons and their postsynaptic targets, granule cells. Quantitative features of this circuit, such as the number of presynaptic neurons that contact a postsynaptic granule cell, have thus far been largely inaccessible. Although it is often assumed that granule cells integrate signals from converging inputs arising from different cells<sup>20</sup>, this has not been shown directly. Multiple presynaptic terminals of one mossy fibre axon are sometimes densely clustered in small regions of the IGL<sup>21</sup>, raising the possibility that a single mossy fibre provides multiple inputs to the same granule cell.

**Reconstruction.** We used confocal microscopy to image the cerebellar flocculus from an adult (P79) *CAGGS-CreERT2<sup>+/+</sup>*; *Thy1-Brainbow-1.0<sup>+/+</sup>* mouse (line H) treated with tamoxifen at P0 (Fig. 5a). Spectrally non-overlapping signals from each of the three XFPs were collected into separate channels (R, G and B). We observed a wide range of colour combinations in different axons, allowing us to distinguish neighbouring axons and trace them through the volume by computer-assisted methods (Fig. 5b and Supplementary Movie 1). The colour information markedly facilitated the tracing process (Supplementary Fig. 4). In total, we reconstructed 341 axons and 93 granule cells. These data included 236 mossy fibre presynaptic terminals or 'rosettes' (Fig. 5c).

Potentially, colour might be used to verify the identity of all the processes arising from an individual neuron without necessarily tracing back to proximal branch points or the soma. This approach would require that the colour profile of a neuronal process remain constant over long distances. In order to study colour constancy, we sampled consecutive mossy fibre rosettes along individual axons and compared their RGB values (Supplementary Fig. 5a, b). The colour profiles obtained for distant regions of a mossy fibre axon (rosettes more than 100  $\mu\text{m}$  apart along a given axon) were largely similar (Fig. 5d). Moreover, axons and dendrites belonging to the same neuron also exhibited similar colour profiles (Supplementary Fig. 6).

Given the colour constancy within a cell, colour differences provide a way to distinguish between neurons and thus could be useful for detailed circuit analysis, such as to count the number of neurons that innervate a postsynaptic cell. Our analysis of cerebellum revealed numerous synaptic interactions between mossy fibres and granule cell dendrites, identified by their characteristic claw-like morphology<sup>22</sup>. We found that individual granule cells were typically innervated by multiple axons that expressed different colours (Fig. 5e, f). Hence, more than one presynaptic neuron innervated each postsynaptic neuron. In several instances, each dendrite of a single granule cell was unambiguously innervated by a different presynaptic neuron (Fig. 5f). In one case, a granule cell was contacted by two presynaptic terminals from a single mossy fibre (Fig. 5g), but it also received inputs from at least one additional (unlabelled) mossy fibre. These data are consistent with the idea that cerebellar granule cells are polyneuronally innervated by mossy fibres.

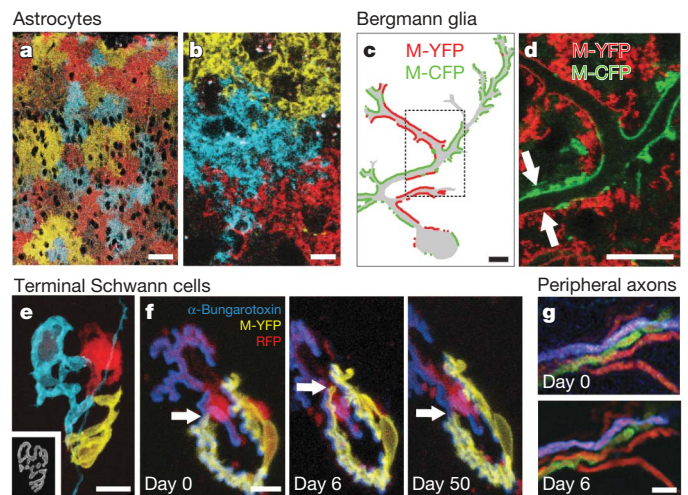
**How many colours?** The usefulness of the Brainbow system to analyse complex connectivity depends on the number of distinguishable colours expressed by neurons. To determine this number, we analysed the distribution of colour profiles in the reconstructed volume from line H above (eight transgene copies). The population of axons exhibited many different colour profiles (Fig. 5c); the mean colour values calculated for the different axons varied greatly in hue and saturation and filled a large portion of colour space (Supplementary Fig. 5c). Using a visual colour discrimination test, we found that

98.9% of randomly selected rosette pairs expressed colours distinct enough to discriminate (see Methods). This degree of colour variation is equivalent to having approximately 89 distinct colours (that is, if 98.9% of axon pairs appear different, then the remaining 1.1% or 1 out of 88.7 pairs are too similar to discriminate). An alternative computer-based colour analysis of hippocampal neuron cell bodies from *Brainbow-1.0* line L (see Fig. 4c) gave an estimated 166 colours. This large number of colours should be useful in resolving individual components of many neural circuits.

### Brainbow analysis of glial interactions

The Brainbow strategy can be used to reveal interactions among neighbouring non-neuronal cells as well as neurons. Glial cells were labelled in four of the *Thy1-Brainbow* lines that we generated (Fig. 6). Two lines showed transgene expression in large populations of astrocytes (Fig. 6a, b; see also Fig. 3d), a third in cerebellar Bergmann glia (Fig. 6c, d) and a fourth in non-myelinating Schwann cells, including those at neuromuscular junctions (Fig. 6e, f).

**Tiling.** The anatomical relations among glial cells and their interactions with neurons have been studied extensively<sup>23,24</sup>. Brainbow provides a new method for approaching this issue. Recombination of *Thy1-Brainbow-1.1* line M led to mutually exclusive expression of three membrane-targeted XFPs in astrocytes throughout the brain, providing unambiguous delineation of boundaries between adjacent astrocytes (Fig. 6a). Single confocal sections showed that adjacent astrocytes were interdigitated in a complex way at their periphery (Fig. 6b). However, each astrocyte possessed a core contiguous territory where it was the only astrocyte present, supporting previous conclusions<sup>23</sup>. Such an exclusive arrangement between individual glial cells and a region of neuropil was not found for a second type of astrocyte labelled in *Thy1-Brainbow-1.0* line G: cerebellar Bergmann glial cells. The thin glial sheath made by Bergmann glia around Purkinje cell dendrites showed extensive interdigitation, with multiple glial cells ensheathing a single portion of dendrite (Fig. 6c, d).



**Figure 6 | Brainbow expression in glial cells and time-lapse imaging.** **a, b**, *Thy1-Brainbow-1.1* line M: astrocytes labelled with three different M-XFPs tile the brain (cortex, **a**) and interdigitate (colliculus, **b**). **c, d**, *Thy1-Brainbow-1.0<sup>(Flp)</sup>* line G: M-YFP and M-CFP (shown as red and green) are expressed in a mutually exclusive manner in Bergmann glial cells of the cerebellum (RFP channel not shown). Reconstruction of a Purkinje cell (grey) reveals tiled Bergmann glial cells associated with different parts of its dendritic tree (**c**) and tightly apposed (**d**, arrows). **e**, *Thy1-Brainbow-1.0* line B: three non-myelinating Schwann cells (blue, yellow, red) tile one neuromuscular junction (inset: cholinergic receptors labelled with Alexa647-bungarotoxin). **f**, Time-lapse images show changes in the territory of individual Schwann cells over 50 days. The cell labelled with M-YFP extends and then retracts a process (arrow). **g**, Time-lapse images of a neuromuscular nerve. The colour of individual axons remained constant over the period imaged. Scale bars: **a**, 50  $\mu\text{m}$ ; **b–g**, 10  $\mu\text{m}$ .

Thus, Bergmann glial cells do not appear to occupy domains related to specific parts of neurons, nor to have exclusive control of an uninterrupted territory.

### Time-lapse imaging

At the adult neuromuscular junction in *Thy1-Brainbow-1.0* line B, it was sometimes possible to visualize each of the individual terminal Schwann cells (Fig. 6e). We performed time-lapse studies on these cells. In 15 out of 17 junctions (in seven mice) imaged over intervals of 4–50 days, we found that the boundaries between adjacent glial cells ensheathing the same motor axon terminal shifted with time, sometimes alternating back and forth (Fig. 6f).

In long-term time-lapse studies of axons that involve repetitive imaging, Brainbow might be useful to identify a specific neuron in successive imaging sessions, assuming a neuron retains its colour. To explore this possibility we imaged the same region of a neuromuscular nerve over 6 days and found that *Brainbow* expression in individual axons remained constant over time both in terms of colour and intensity (Fig. 6g). Thus, the colour tags provided by *Brainbow* are stable and expressed at sufficient levels to permit longitudinal studies of circuits *in vivo*.

### Discussion

*Brainbow* transgenes are novel reagents for large-scale studies of cellular interactions. These Cre/*lox*-based transgenes create a mosaic gene expression offering two key advantages. First, they use DNA excision and/or inversion to create a stochastic choice among several genes—up to four with the configurations presented here as opposed to two with other methods<sup>9–11</sup>. Notably, no specific pairs of *lox* sites or transgene orientations are strongly favoured, so distinct outcomes can occur at similar frequencies. Moreover, following recombination the expression choice is stabilized, either by using *lox* variants which prevent further excision in *Brainbow-1*, or by transient Cre action (permitted by several strategies<sup>11,15,25</sup>) to stop inversion in *Brainbow-2*. Second, this stochastic choice can give rise to either mutually exclusive gene expression when a single copy of the construct is present, or to combinatorial expression when there are multiple copies. Several parameters might affect the diversity of the combinations obtained, including promoter choice, transgene copy number and length, efficiency and duration of the recombination.

In the constructs presented here, combinatorial expression of three XFP genes generates approximately 100 colours in neurons. This labelling appears well suited for visualization and tracing of large numbers of neurons and their connectivity. We show that colour differences between neurons provide a way to sort their processes while tracing through sections, to directly visualize their putative synaptic interactions, and to distinguish the neurons that converge onto a postsynaptic cell. The power of this circuit analysis is presently limited by the number and distribution of colours and the ability to resolve them; the lack of a concomitant way to identify unambiguously synaptic contacts; and the restriction of *Thy1* expression to certain neuronal populations. Future constructs could address most of these limitations by taking advantage of the ever-expanding spectrum of fluorescent proteins<sup>5</sup>, promoters<sup>26,27</sup> and Cre drivers<sup>10</sup> as well as by combining Brainbow with transgenic synaptic markers or ultra-thin sectioning techniques<sup>3</sup>.

Brainbow's ability to switch on multiple genes may also prove useful for other purposes. The *Brainbow* cassettes could be adapted to any cell type and to any organism in which site-specific recombinases act. We anticipate that Brainbow could be used in lineage analysis to mark neighbouring clones with distinct colours. Brainbow will also be useful to visualize individual cells and their interactions, as demonstrated here for both neurons and glial cells. Finally, this approach raises tantalizing possibilities for modulating gene expression and comparing the effects of several distinct genetic changes (or their combination) simultaneously introduced into a population of cells.

### METHODS SUMMARY

*Brainbow* constructs were assembled by standard cloning methods using the following XFPs: mEYFP (monomeric EYFP<sup>28</sup>), mCerulean<sup>29</sup>, Kusabira orange<sup>30</sup>, tdimer2(12)<sup>31</sup>, mCherry<sup>32</sup>, dTomato<sup>32</sup>, ECFP and dsRed2 (Clontech), and hrGFP<sup>II</sup>-NLS (Stratagene). A palmitoylation sequence<sup>33</sup> tethered XFPs to the membrane in some constructs, allowing clear labelling of axonal processes, whereas cytoplasmic XFPs better labelled neuronal cell bodies and dendrites. A novel *lox* variant, *loxN*, was created with mutations distinct from both *loxP* and *lox2272*, making all three *lox* sites mutually incompatible. For *in vitro* tests, HEK cells stably expressing *CMV-Brainbow* constructs were selected. Recombination was induced by transient *CMV-Cre* transfection. For transgenic mouse generation, *Brainbow* constructs were cloned in a *Thy1.2* cassette<sup>13</sup>. Three to nine independent lines were obtained for each construct and crossed with *CAGGS-CreERT2* mice<sup>14</sup>. Their offspring received a single injection of 10–50 µg tamoxifen between P0 and P3 (ref. 34). For transgene copy reduction, mice were crossed with  $\beta$ -actin-*Flpe* animals<sup>18</sup>. Anaesthetized mice were perfused with 2–4% PFA. Brain sections, whole-mount skeletal muscle, nerves and fixed HEK cells were imaged using a Zeiss Imager.Z1 epifluorescence microscope (with standard CFP, YFP, TRITC and TxRed filters), or an Olympus FV1000 confocal microscope (equipped with 440-, 515- and 568-nm lasers). Live imaging methods were adapted from ref. 35. In all lines XFP fluorescence was directly observable, except for *Brainbow-2.0* animals, which required anti-GFP immunostaining. Linear unmixing was used only for the constructs expressing overlapping dyes (OFP/RFP and CFP/GFP/YFP). Tracing and three-dimensional reconstruction were performed using Reconstruct software<sup>36</sup>. Colour profiles of axon sections were analysed and displayed using Matlab. Visual- and computer-based colour quantifications are described in Methods.

Four of the most generally useful *Thy1-Brainbow* mouse lines described here (H, L, M and R) are available to the academic research community through Jackson Laboratory (JAX Stock Numbers 007901, 007910, 007911 and 007921, respectively).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 17 July; accepted 25 September 2007.**

- Denk, W. & Horstmann, H. Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS Biol.* **2**, e329 (2004).
- Briggman, K. L. & Denk, W. Towards neural circuit reconstruction with volume electron microscopy techniques. *Curr. Opin. Neurobiol.* **16**, 562–570 (2006).
- Micheva, K. D. & Smith, S. J. Array tomography: a new tool for imaging the molecular architecture and ultrastructure of neural circuits. *Neuron* **55**, 25–36 (2007).
- Gan, W. B., Grutzendler, J., Wong, W. T., Wong, R. O. & Lichtman, J. W. Multicolor “DiOlistic” labeling of the nervous system using lipophilic dye combinations. *Neuron* **27**, 219–225 (2000).
- Shaner, N. C., Steinbach, P. A. & Tsien, R. Y. A guide to choosing fluorescent proteins. *Nature Methods* **2**, 905–909 (2005).
- Hadjantonakis, A. K., Macmaster, S. & Nagy, A. Embryonic stem cells and mice expressing different GFP variants for multiple non-invasive reporter usage within a single animal. *BMC Biotechnol.* **2**, 11 (2002).
- Feng, G. et al. Imaging neuronal subsets in transgenic mice expressing multiple spectral variants of GFP. *Neuron* **28**, 41–51 (2000).
- Kasthuri, N. & Lichtman, J. W. The role of neuronal identity in synaptic competition. *Nature* **424**, 426–430 (2003).
- Zong, H., Espinosa, J. S., Su, H. H., Muzumdar, M. D. & Luo, L. Mosaic analysis with double markers in mice. *Cell* **121**, 479–492 (2005).
- Branda, C. S. & Dymecki, S. M. Talking about a revolution: The impact of site-specific recombinases on genetic analyses in mice. *Dev. Cell* **6**, 7–28 (2004).
- Lam, K. P. & Rajewsky, K. Rapid elimination of mature autoreactive B cells demonstrated by Cre-induced change in B cell antigen receptor specificity *in vivo*. *Proc. Natl Acad. Sci. USA* **95**, 13171–13175 (1998).
- Lee, G. & Saito, I. Role of nucleotide sequences of *loxP* spacer region in Cre-mediated recombination. *Gene* **216**, 55–65 (1998).
- Caroni, P. Overexpression of growth-associated proteins in the neurons of adult transgenic mice. *J. Neurosci. Methods* **71**, 3–9 (1997).
- Guo, C., Yang, W. & Lobe, C. G. A Cre recombinase transgene with mosaic, widespread tamoxifen-inducible action. *Genesis* **32**, 8–18 (2002).
- Metzger, D., Clifford, J., Chiba, H. & Chambon, P. Conditional site-specific recombination in mammalian cells using a ligand-dependent chimeric Cre recombinase. *Proc. Natl Acad. Sci. USA* **92**, 6991–6995 (1995).
- Garrick, D., Fiering, S., Martin, D. I. & Whitelaw, E. Repeat-induced gene silencing in mammals. *Nature Genet.* **18**, 56–59 (1998).
- Palmiter, R. D. & Brinster, R. L. Germ-line transformation of mice. *Annu. Rev. Genet.* **20**, 465–499 (1986).
- Rodriguez, C. I. & Dymecki, S. M. Origin of the precerebellar system. *Neuron* **27**, 475–486 (2000).



19. Rowan, S. & Cepko, C. L. Genetic analysis of the homeodomain transcription factor Chx10 in the retina using a novel multifunctional BAC transgenic mouse reporter. *Dev. Biol.* **271**, 388–402 (2004).
20. Eccles, J. C., Ito, M. & Szentágothai, J. in *The Cerebellum as a Neuronal Machine* (Springer, Berlin, 1967).
21. Wu, H. S., Sugihara, I. & Shinoda, Y. Projection patterns of single mossy fibers originating from the lateral reticular nucleus in the rat cerebellar cortex and nuclei. *J. Comp. Neurol.* **411**, 97–118 (1999).
22. Palay, S. L. & Chan-Palay, V. in *Cerebellar Cortex* (Springer, New York, 1974).
23. Bushong, E. A., Martone, M. E., Jones, Y. Z. & Ellisman, M. H. Protoplasmic astrocytes in CA1 stratum radiatum occupy separate anatomical domains. *J. Neurosci.* **22**, 183–192 (2002).
24. Ogata, K. & Kosaka, T. Structural and quantitative analysis of astrocytes in the mouse hippocampus. *Neuroscience* **113**, 221–233 (2002).
25. Silver, D. P. & Livingston, D. M. Self-excising retroviral vectors encoding the Cre recombinase overcome Cre-mediated cellular toxicity. *Mol. Cell* **8**, 233–243 (2001).
26. Gong, S. *et al.* A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature* **425**, 917–925 (2003).
27. Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).
28. Zacharias, D. A., Violin, J. D., Newton, A. C. & Tsien, R. Y. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science* **296**, 913–916 (2002).
29. Rizzo, M. A., Springer, G. H., Granada, B. & Piston, D. W. An improved cyan fluorescent protein variant useful for FRET. *Nature Biotechnol.* **22**, 445–449 (2004).
30. Karasawa, S., Araki, T., Nagai, T., Mizuno, H. & Miyawaki, A. Cyan-emitting and orange-emitting fluorescent proteins as a donor/acceptor pair for fluorescence resonance energy transfer. *Biochem. J.* **381**, 307–312 (2004).
31. Campbell, R. E. *et al.* A monomeric red fluorescent protein. *Proc. Natl Acad. Sci. USA* **99**, 7877–7882 (2002).
32. Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nature Biotechnol.* **22**, 1567–1572 (2004).
33. Kay, J. N. *et al.* Transient requirement for ganglion cells during assembly of retinal synaptic layers. *Development* **131**, 1331–1342 (2004).
34. Buffelli, M. *et al.* Genetic evidence that relative synaptic efficacy biases the outcome of synaptic competition. *Nature* **424**, 430–434 (2003).
35. Zuo, Y. *et al.* Fluorescent proteins expressed in mouse transgenic lines mark subsets of glia, neurons, macrophages, and dendritic cells for vital examination. *J. Neurosci.* **24**, 10999–11009 (2004).
36. Fiala, J. C. Reconstruct: a free editor for serial section microscopy. *J. Microsc.* **218**, 52–61 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank members of the Lichtman and Sanes laboratories for discussions; H. Nishimune for insights; A. Mendelsohn and A. Bruskin for cell culture and PCR; S.-H. Shieu and D. Mou for imaging software development; S. Haddad, D. Pelusi, D. Malkowski, K. Mahoney and the Harvard MCB BRI facility for animal care and genotyping; M. Wallace, the WUSTL Mouse Genetic Core and the MCB Genome Manipulation Facility for pronuclei injections; and R. Hellmiss for help with graphics. We thank J. C. Fiala for assistance with Reconstruct; M. L. Nonet, J. S. Mumm and R. O. Wong for advice and reagents; D. W. Piston for mCerulean; R. Y. Tsien for tdimer2 and mCherry; and A. Miyawaki for Kusabira. This work was supported by the James S. McDonnell foundation and grants from NIH.

**Author Contributions** J.L., J.R.S. and J.W.L. conceived the Brainbow strategies. J.R.S. and J.W.L. supervised the project. J.L. built initial constructs and validated them *in vitro* and *in vivo*. T.A.W. performed all cerebellar axonal tracing and colour profile analysis with programs developed with J. Lu. H.K. performed all live imaging experiments. R.W.D. generated *Brainbow-1.0* lines expressing cytoplasmic XFPs, and R.A.B. generated *Brainbow-1.1* constructs and lines. J.L., T.A.W. and R.W.D. screened mouse lines.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to J.W.L. ([jeff@mcb.harvard.edu](mailto:jeff@mcb.harvard.edu)).

## METHODS

**DNA constructs.** *Brainbow* constructs were assembled by standard cloning methods, using the CMV expression vector pEYFP-N1 (Clontech) as a start point. Briefly, annealed oligonucleotides were used to insert *lox* sites in pEYFP-N1 to generate *lox-EYFP*. Other fluorescent proteins were then cloned in place of *EYFP*, and these *lox-XFP* units were ligated stepwise to generate the *CMV-Brainbow* constructs used for *in vitro* tests. For the *Brainbow-2.1* construct, a nuclear localization signal (NLS) was added to the carboxy-terminal end of hrGFP11 to permit unambiguous distinction from its close spectral neighbours CFP and YFP. For the *Brainbow-1.1* construct, a novel *lox* variant, *loxN*, was created with substitutions in position 2 and 7 of the *lox* spacer that differed from both *loxP* and *lox2272*, making all three *lox* sites mutually incompatible (Supplementary Table 1). An *FRT* site was inserted at the 3' end of each *Brainbow* construct to permit reduction of transgene multicopy by action of F1p recombinase (Fig. 4d and ref. 16). *Brainbow* cassettes were subsequently cloned in the *XhoI* site of a *Thy1.2* genomic element<sup>13</sup>. Except in *Brainbow-2.0*, each reading frame was followed by a polyadenylation signal (SV40, bGH, or from the 3' part of the *Thy1* element) to maximize expression<sup>37</sup>.

**Cell culture.** *CMV-Brainbow* constructs bearing a *neo<sup>r</sup>* selection marker were linearized and electroporated into HEK cells. Stable transfectants were selected with G418 and fluorescent clones were picked and expanded. Recombination was induced in these cells by transient transfection with a *CMV-Cre* vector using Lipofectamine 2000 (Invitrogen). After 3–12 days *in vitro*, cells were fixed for 10 min with cold 4% PFA, and samples were mounted in Vectashield before imaging.

**Mice.** *Thy1-Brainbow* constructs were linearized with *EcoRI* and *PvuII* and injected into pronuclei to generate 3–9 independent transgenic lines per construct. Genotyping and transgene copy number estimation were performed with PCR using YFP primers. Lines were crossed with CAGGS-*CreERT2* mice<sup>14</sup>. Their offspring received a single dorsal subcutaneous injection of 10–50 µg tamoxifen (Sigma T5648, 1–5 mg ml<sup>-1</sup> in corn oil) between P0 and P3, in order to induce widespread recombination throughout the nervous system<sup>34</sup>. *Thy1-Brainbow-1.0* lines F and H were also crossed to four *Cre* driver lines known to trigger recombination in specific nervous system regions: *Isl1<sup>Cre</sup>* in retina and basal ganglia<sup>38</sup>, *Chx10-Cre* and *Pax6<sub>ex</sub>-Cre* in retina<sup>19,39</sup>, and *Emx1<sup>Cre</sup>* in cortex<sup>40</sup>. In each case, we observed *Brainbow* expression (YFP, CFP and combinations) in the territories where expression of *Thy1* and the *Cre* driver intersect (Fig. 3e; Supplementary Fig. 3 and data not shown). For germline F1p-induced transgene copy reduction, mice were mated to  $\beta$ -*actin-Flpe* animals<sup>18</sup>. Interestingly, transgene copy reduction of *Brainbow-1.0* line H by F1p/*FRT* necessitated three generations before a line showing mutually exclusive XFP expression was derived. This may be related to the large size of the construct (10 kb for *Thy1-Brainbow-1.0*), as site-specific recombination is known to be affected by distance<sup>41,42</sup>. To verify transgene copy number reduction, we used a PCR spanning the junction between adjacent transgene copies.

Four of the most generally useful *Thy1-Brainbow* mouse lines described here (H, L, M and R) are available to the academic research community through Jackson Laboratory (JAX Stock 007901, 007910, 007911 and 007921, respectively).

**Tissue preparation.** Animals between P8 and P40 were anaesthetized with sodium pentobarbital before intracardiac perfusion with PBS and 4% PFA. One-hundred-micrometre sagittal brain sections were obtained using a vibrating microtome (Leica VT1000). Sections and fixed tissues were mounted in Vectashield and stored at -20 °C before imaging. In the case of *Brainbow-2.0* mice, which possessed the only construct that did not include polyadenylation signals directly after each XFP, the CFP signal was amplified with an anti-GFP antibody (Chemicon Ab3080).

**Imaging.** Fixed brain and muscle samples were imaged with  $\times 20$  (0.8 NA),  $\times 40$  (1.3 NA) or  $\times 60$  (1.45 NA) oil objectives. Confocal images were acquired with an Olympus FV1000 microscope equipped with a 440/515/568/633-nm primary

beam splitter, using a 440-nm photodiode laser for CFP (PMT1, 510-nm dichroic mirror, 465–495-nm barrier filter), a 515-nm argon line for YFP (PMT2, 560-nm dichroic mirror, 530–565-nm barrier filter), and a 561-nm photodiode laser or 568-nm krypton laser for RFP (PMT3, 585–615-nm barrier filter). Confocal image stacks for all three channels were acquired sequentially, and maximally projected using Metamorph or ImageJ software. For cell culture experiments, linear unmixing was performed with Olympus or ImageJ software on single-plane confocal images obtained on the setup described above to separate OFP/RFP signals in *Brainbow-1.1* and CFP/GFP/YFP signals in *Brainbow-2.0*, using a 10-nm collection window and 10-nm step for the lambda scan (in some cases linear unmixing was performed on epifluorescence images of cell cultures).

In Adobe Photoshop, the 12-bit images obtained for each channel were imported in the 16-bit mode and then mapped to an 8-bit display before uniform adjustment of levels, contrast, brightness and gamma. The channels were then assigned colours and overlaid. Unless noted, CFP, YFP and RFP fluorescence was displayed using cyan, yellow and red colour (in some cases using blue, green and red).

**Tracing and three-dimensional reconstruction.** We used Reconstruct, a program to assist three-dimensional reconstruction from serial sections available at <http://synapses.bu.edu><sup>36</sup>. Confocal data stacks were loaded as RGB tiffs in either the XY or YZ plane. Tracing of cross-sections was done using the automated 'wildfire' tool where possible, otherwise by hand.

**Colour analysis.** We used Matlab software to measure colour in each traced section throughout the data set, then calculated the pixel distribution and centroid (or mean colour) for each reconstructed axon. The Matlab *rgb2hsv* function was used to translate centroid colours from RGB to HSV coordinates. In the visual colour-discrimination test (done on mossy fibre rosettes from *Brainbow-1.0* line H), we compared pairs of images randomly selected from the 236 reconstructed mossy fibre rosettes ( $n = 6,225$ ). Matlab presented random samples from two rosettes on the computer screen and an observer determined whether the samples were drawn from the same or different rosettes. In the computer-based analysis (done on dentate gyrus granule cells from *Brainbow-1.0* line L), a colour criterion was calculated (eight units in RGB space) based on a within-cell analysis that compared the colour centroids of two regions within each cell soma. This criterion was then used for a between-cell comparison that searched all the other imaged neurons ( $>200$ ) to determine the total number of distinct colours present in the data set. The result showed that there were 166 different colours (128 neurons expressed unique colours, 52 were indistinguishable from one other cell, 20 were indistinguishable from two other cells and two were indistinguishable from three other cells).

**Live imaging.** Methods for repetitive vital imaging of 1–3-month-old mice were adapted from ref. 35. Acetylcholine receptors were labelled by application of Alexa-647-conjugated  $\alpha$ -bungarotoxin (Molecular Probes) in sterile, lactated Ringer's solution. We used the FV1000 confocal setup described above equipped with a  $\times 20$ , 1.0 NA water immersion objective. All channels were acquired simultaneously.

37. Kakoki, M. *et al.* Altering the expression in mice of genes by modifying their 3' regions. *Dev. Cell* **6**, 597–606 (2004).
38. Srinivas, S. *et al.* Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev. Biol.* **1**, 4 (2001).
39. Marquardt, T. *et al.* Pax6 is required for the multipotent state of retinal progenitor cells. *Cell* **105**, 43–55 (2001).
40. Gorski, J. A. *et al.* Cortical excitatory neurons and glia, but not GABAergic neurons, are produced in the *Emx1*-expressing lineage. *J. Neurosci.* **22**, 6309–6314 (2002).
41. Ringrose, L., Chabanis, S., Angrand, P. O., Woodroffe, C. & Stewart, A. F. Quantitative comparison of DNA looping *in vitro* and *in vivo*: chromatin increases effective DNA flexibility at short distances. *EMBO J.* **18**, 6630–6641 (1999).
42. Zheng, B., Sage, M., Sheppard, E. A., Jurecic, V. & Bradley, A. Engineering mouse chromosomes with Cre-loxP: range, efficiency, and somatic applications. *Mol. Cell. Biol.* **20**, 648–655 (2000).



# Dissecting a circuit for olfactory behaviour in *Caenorhabditis elegans*

Sreekanth H. Chalasani<sup>1</sup>, Nikos Chronis<sup>1</sup>, Makoto Tsunozaki<sup>1</sup>, Jesse M. Gray<sup>1</sup>, Daniel Ramot<sup>2</sup>, Miriam B. Goodman<sup>2</sup> & Cornelia I. Bargmann<sup>1</sup>

Although many properties of the nervous system are shared among animals and systems, it is not known whether different neuronal circuits use common strategies to guide behaviour. Here we characterize information processing by *Caenorhabditis elegans* olfactory neurons (AWC) and interneurons (AIB and AIY) that control food- and odour-evoked behaviours. Using calcium imaging and mutations that affect specific neuronal connections, we show that AWC neurons are activated by odour removal and activate the AIB interneurons through AMPA-type glutamate receptors. The level of calcium in AIB interneurons is elevated for several minutes after odour removal, a neuronal correlate to the prolonged behavioural response to odour withdrawal. The AWC neuron inhibits AIY interneurons through glutamate-gated chloride channels; odour presentation relieves this inhibition and results in activation of AIY interneurons. The opposite regulation of AIY and AIB interneurons generates a coordinated behavioural response. Information processing by this circuit resembles information flow from vertebrate photoreceptors to 'OFF' bipolar and 'ON' bipolar neurons, indicating a conserved or convergent strategy for sensory information processing.

Neural circuits actively transform sensory signals: they extract the most relevant sensory information from the environment, determine whether stimuli are increasing or decreasing, and use this information to regulate behaviours on timescales from seconds to hours. It has been suggested that a few connected neurons in statistically over-represented synaptic 'motifs' could perform simple circuit computations<sup>1</sup>. The potential to analyse circuit function at single-cell resolution exists in the nematode *C. elegans*, the nervous system of which contains just 302 neurons with known synaptic connections<sup>2–4</sup>. Functions for many individual *C. elegans* neurons have been inferred from cell ablation studies, chronic activation or recording of neuronal activity<sup>5–10</sup>. However, little has been done to study the dynamic flow of information between neurons—an essential link between circuits and behaviour.

In some *C. elegans* behaviours, sensory inputs lead to long-lasting and complex behavioural sequences. During chemotaxis to food odours or tastes, animals reach an attractant source by regulating turns over many minutes<sup>5,11</sup>. In another food-related behaviour called local search or area-restricted search, animals that have recently been removed from food spend about 15 min exploring a restricted area by interrupting long forward movements with stochastic turns, and then disperse by suppressing turning<sup>12–15</sup>. A pair of olfactory neurons called AWC is important in both of these food-seeking behaviours: the AWC neurons direct chemotaxis to many attractive odours<sup>5</sup>, and also increase turning probability during local search<sup>14</sup>. The AWC neurons synapse onto several interneurons including AIB and AIY, which enhance and suppress turning, respectively<sup>13,14</sup> (Fig. 1a). Although the sensory signalling of the AWC neuron has been extensively characterized at a genetic level, nothing is known about the effects of odours on AWC neuron activity, or the mechanisms by which the AWC neuron communicates with downstream neurons. Here we analyse the functional connectivity between the AWC, AIB and AIY neurons that initiates the transformation of chemosensory cues into behaviour.

## AWC neurons respond to odour removal

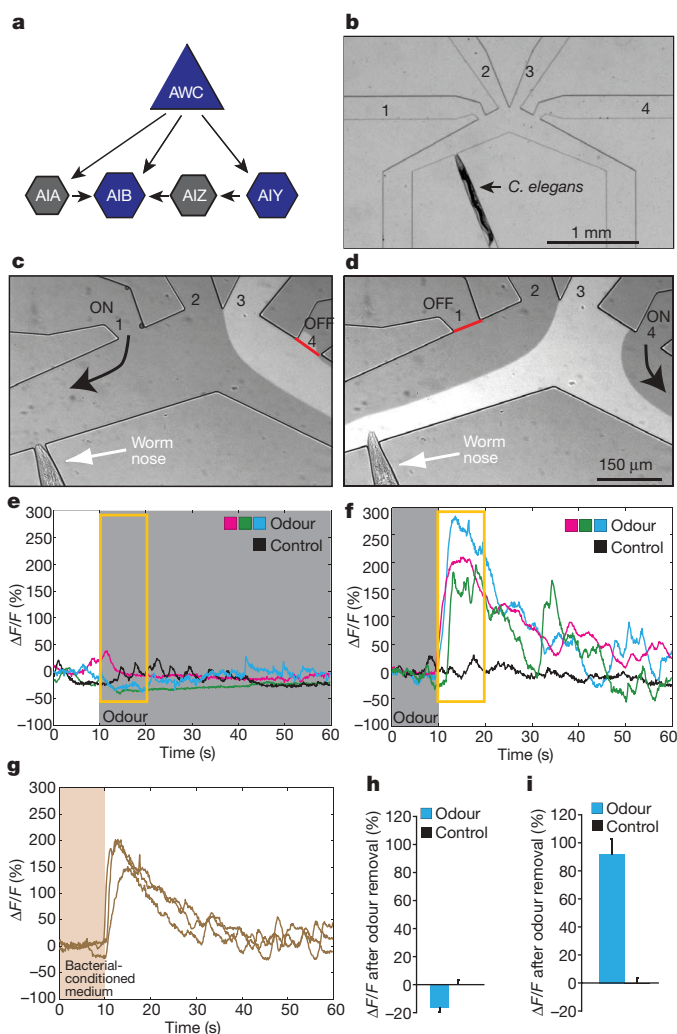
To monitor the odour response of AWC neurons, we used AWC-selective promoters to express G-CaMP, a genetically encoded calcium sensor, the fluorescence intensity of which increases on calcium binding in the physiological range<sup>16,17</sup>. *C. elegans* neurons have rapidly activating voltage-gated calcium channels, suggesting that calcium transients should correlate with strong neuronal depolarization<sup>9,18</sup>. A custom-designed microfluidic device fabricated from the transparent polymer polydimethylsiloxane was used to trap and stimulate animals expressing G-CaMP in one of the two AWC neurons, AWC<sup>ON</sup> (Fig. 1b–d)<sup>19</sup>. The device restrained z-plane movement of the animal so that the AWC neuron remained in focus for imaging, while allowing limited movement in the x–y plane. Fluid streams under laminar flow were used to deliver the buffer or odour stimuli to the animal's nose. Buffer exchange controls ( $n = 18$ ) showed that short-motion transients, photobleaching and other measurement errors contributed  $\Delta F/F$  (fluorescence change/baseline fluorescence intensity) signals <25% to individual traces, and <10% to averaged traces (Supplementary Figs 1–3); averaged traces were used for quantification and statistical analysis (Supplementary Fig. 2).

Individual animals were followed in a paired odour addition–odour removal sequence (Fig. 1e, f). Presentation of the attractive odour isoamyl alcohol diminished G-CaMP fluorescence in the AWC<sup>ON</sup> cell body; although slight, the decrease was significant when averaged over multiple trials (Fig. 1e, h and Supplementary Figs 1–3). Large increases in G-CaMP fluorescence were reliably observed within one second when isoamyl alcohol was replaced with buffer, suggesting an increase in AWC<sup>ON</sup> calcium on odour removal (an odour-OFF response) (Fig. 1f, i, Supplementary Fig. 1b and Supplementary Movie 1). Increases in G-CaMP fluorescence in AWC<sup>ON</sup> were also observed when bacteria-conditioned medium was removed from the animal's nose (Fig. 1g and Supplementary Fig. 1h). The AWC<sup>ON</sup> odour-OFF response was observed at isoamyl alcohol dilutions of  $10^{-7}$  to  $10^{-3}$ , with a maximal response at  $10^{-4}$ .

<sup>1</sup>Howard Hughes Medical Institute, Laboratory of Neural Circuits and Behaviour, The Rockefeller University, New York, New York 10065, USA. <sup>2</sup>Program in Neurobiology and Department of Molecular and Cellular Physiology, Stanford University, Stanford, California 94305, USA.

(Fig. 2a); inhibition by odour was maximal at  $10^{-5}$  to  $10^{-4}$  dilution (Supplementary Figs 2 and 3). Chemotaxis to isoamyl alcohol saturates when the odour is uniformly added to agar at  $\sim 10^{-4}$  dilution, approximately matching the saturation of the AWC<sup>ON</sup> calcium response<sup>20,21</sup>.

Single *C. elegans* olfactory neurons express multiple G-protein-coupled receptors and sense multiple odours<sup>5</sup>. The AWC<sup>ON</sup> neuron senses benzaldehyde as well as isoamyl alcohol, and an odour-OFF response was observed in AWC<sup>ON</sup> across  $10^{-6}$  to  $10^{-4}$  dilutions of benzaldehyde (Fig. 2b). The contralateral AWC<sup>OFF</sup> neuron also showed odour-OFF G-CaMP responses to isoamyl alcohol and benzaldehyde—a result consistent with behavioural evidence that both AWC<sup>ON</sup> and AWC<sup>OFF</sup> sense these two attractive odours<sup>22</sup> (Fig. 2c, d). At low dilutions, the odour 2,3-pentanedione is sensed by AWC<sup>OFF</sup> and not by AWC<sup>ON</sup> in behavioural assays, and similarly AWC<sup>OFF</sup> responded to 2,3-pentanedione at  $10^{-7}$  dilution whereas AWC<sup>ON</sup> did not (Fig. 2e). In all cases, odour removal increased G-CaMP fluorescence, and odour addition diminished fluorescence.



**Figure 1 | AWC responds to odour removal.** **a**, AWC sensory neurons and downstream interneurons. **b**, Low magnification and **c**, **d**, high magnification view of the PDMS imaging chip, with worm nose exposed to buffer (**c**, streams 1–3 are open; stream 2 reaches the nose) or odour (**d**, streams 2–4 are open; stream 3 reaches the nose). **e–g**, Representative G-CaMP responses from AWC<sup>ON</sup>. **e**, **f**, AWC<sup>ON</sup> responses on addition (**e**) or removal (**f**) of isoamyl alcohol odour (coloured traces) or buffer (black traces) at 10 s. Grey shading denotes presence of odour; yellow intervals are analysed in **h** and **i**. **g**, AWC<sup>ON</sup> responses on removal of bacterial-conditioned medium at 10 s. **h**, **i**, Average fluorescence change in AWC<sup>ON</sup> during the 10 s after odour addition (**h**) or removal (**i**) ( $n = 30$ ). Error bars indicate standard error of the mean (s.e.m.).

The strong bilateral AWC response to  $10^{-4}$  isoamyl alcohol, which resembled the response to bacterial-conditioned medium, was used for subsequent characterization. The AWC<sup>ON</sup> odour-OFF response was long-lasting, with an average half-time of  $\sim 20$  s and large secondary peaks throughout the 50 s recording period that were not observed in the presence of odour (Supplementary Fig. 1a), in buffer controls (Supplementary Fig. 1c, d) or with calcium-insensitive green-fluorescent protein (Supplementary Fig. 1e, f). Increasing the duration of odour pre-exposure from 1 to 5 min resulted in a graded odour-OFF calcium response, with a stronger response after longer odour exposure (Fig. 2f). Odour-OFF responses were silenced when isoamyl alcohol was added back after 10 s or 30 s (Fig. 2g, h). Removal of isoamyl alcohol 20 s later elicited a second odour-OFF calcium response of smaller magnitude ( $42\text{--}46\% \pm 14\%$  compared to the first response, Fig. 2g, h), suggesting that AWC integrates odour history over  $>20$  s. These results suggest that AWC neurons are inhibited by attractive odours, and depolarize when odours are removed, with a response that integrates the concentration and duration of odour exposure.

### Odour responses in AIB and AIY interneurons

The AIB and AIY interneurons receive synapses from the AWC neurons, and function with the AWC neurons in local search behaviour<sup>3,13,14</sup> (Fig. 1a). AIB and AIY responses to odours were monitored in the microfluidic device using transgenic lines that expressed G-CaMP under cell-selective promoters. Like AWC, the AIB cell body responded strongly to odour removal with increased calcium, and was inhibited by odour addition (Fig. 3a, b, Supplementary Figs 3 and 4a, b and Supplementary Movie 2). AIB responses were eliminated in animals in which the AWC sensory neurons were ablated (Fig. 3b and Supplementary Fig. 4c, d). These results indicate that the AWC neurons activate AIB interneurons on odour removal.

AIB calcium responses were sustained compared to AWC responses, persisting at near-peak levels for at least 2 min after odour removal (Fig. 3b and data not shown). Two possible explanations, not mutually exclusive, could account for this persistent activity: the AIB interneurons might respond to persistent transmitter release from AWC neurons during the late phase of the response, or AIB interneurons might remain active without AWC input through intrinsic persistent activity or reverberant circuit activity<sup>23</sup>. To examine the first possibility, we removed odour to activate AIB interneurons and then provided odour 30 s later during the period of sustained AIB activity. AIB responses decayed after odour addition, suggesting that AWC neurons continued to stimulate AIB interneurons by releasing transmitter after the end of the strong AWC calcium response (Fig. 3c and Supplementary Fig. 4e, f). This conclusion about AWC is consistent with the known property of nematode (*Ascaris*) motor neurons, which release neurotransmitter in a graded fashion<sup>24</sup>.

Like AWC neurons, the AIB interneurons had a graded calcium response to the duration of odour exposure, with longer and stronger responses after 5 min of odour pre-exposure compared to 3 min or 1 min (Fig. 3d).

In AIY neurons, changes in G-CaMP fluorescence were detected in the neurite but not in the cell body; similar observations have been made using temperature stimuli and the cameleon calcium sensor in AIY neurons<sup>25</sup>. In contrast with AWC and AIB neurons, increases in AIY calcium were observed on odour addition and decreases were observed on odour removal (Fig. 3e, f, Supplementary Figs 3 and 5a, b and Supplementary Movie 3). AIY responses were brief, poorly synchronized to odour onset, and appeared sporadically throughout the period of odour presentation (Fig. 3e and Supplementary Fig. 5a). Ablation of the AWC neurons eliminated the AIY response (Fig. 3e, f and Supplementary Fig. 5c, d). These results indicate that AIY is inhibited by AWC, even at the resting level of AWC activity, and that release from this inhibition on odour presentation results in AIY activation.



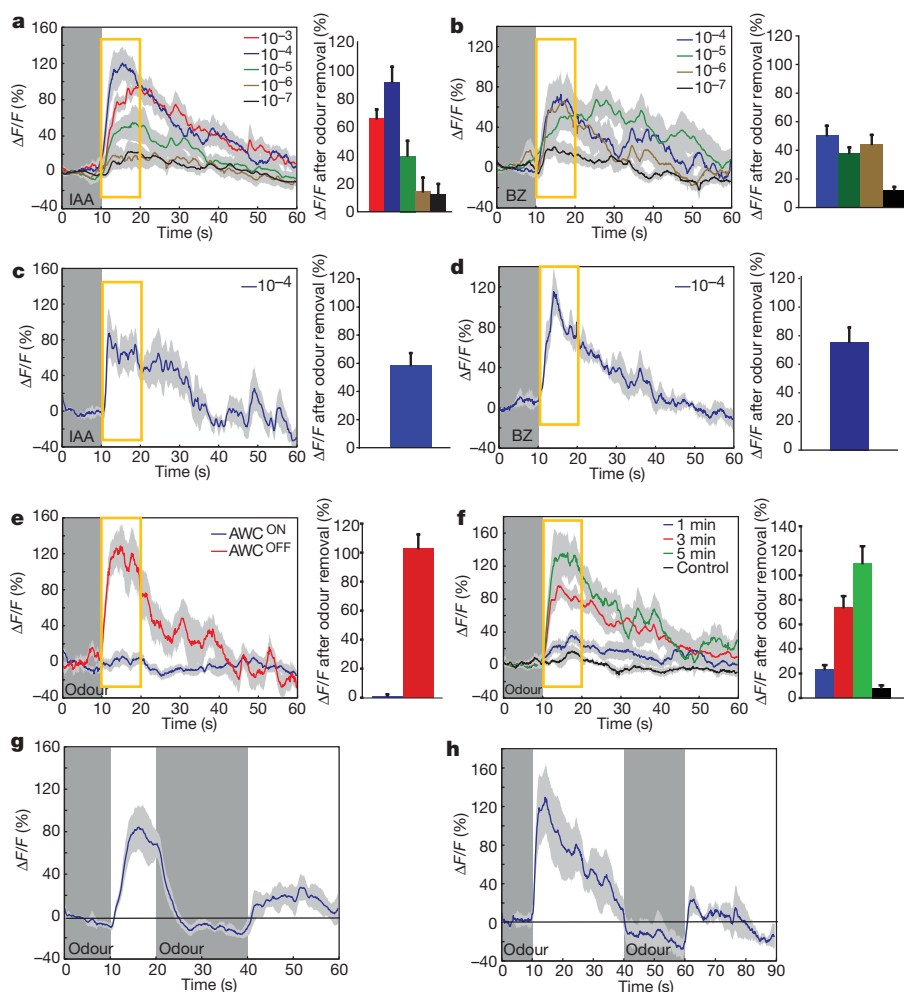
Thus, AWC, AIB and AIY neurons have distinct dynamic responses to odours (Fig. 3g, h). AWC has a transient odour-OFF response that peaks and then decays, and AIB has a sustained odour-OFF response, suggesting that it performs a temporal integration of AWC activity. AIY has an irregular odour-ON response, with short asynchronous events that are not well-represented in averaged traces.

### AWC synapses are glutamate-mediated

AWC neurons have small clear vesicles at their synapses, suggesting the use of a small-molecule neurotransmitter<sup>3</sup>. *C. elegans eat-4(ky5)* mutants are defective in a vesicular glutamate transporter that concentrates glutamate into synaptic vesicles, and are defective in olfactory chemotaxis, local search and numerous other behaviours<sup>12,26</sup>. An *eat-4* complementary DNA expressed from the AWC-selective *odr-3* promoter (*AWC::eat-4*) rescued chemotaxis of *eat-4(ky5)* mutants in odour gradients, suggesting that AWC uses glutamate as its transmitter (Fig. 4a). The transgene also partially restored *eat-4(ky5)* local search behaviour—the AWC-dependent turning behaviour observed in the first 15 min after animals are removed from food (Fig. 4b). In the *AWC::eat-4* transgenic animals, turns were appropriately suppressed in the presence of food and in the later dispersal period, indicating that glutamate from AWC neurons stimulates appropriate local search and not continuous turning

behaviour<sup>6</sup>. These experiments indicate that AWC neurons can release glutamate to stimulate olfactory chemotaxis and local search.

Local search behaviour is a time-dependent, quantitative response to the removal of food odours, so this assay was used for detailed analysis of the circuit downstream of AWC neurons. *C. elegans* has four classes of glutamate receptors: AMPA ( $\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid)-type glutamate-gated cation channels (encoded by eight *glr* genes); NMDA (*N*-methyl-D-aspartate receptor)-type glutamate-gated cation channels (encoded by two *nmr* genes); metabotropic G-protein-coupled glutamate receptors (encoded by two *mgl* genes); and glutamate-gated chloride channels (encoded by 10–20 *glc/avr* genes)<sup>27–30</sup>. The AMPA-type receptor GLR-1 is expressed by AIB interneurons as well as command interneurons that control forward and backward locomotion<sup>27,28</sup>, and *glr-1(n2461)* mutant animals have diminished local search behaviour<sup>12</sup> (Fig. 4c). The local search defect was rescued by expressing *glr-1* from a promoter expressed in AIB and AIZ interneurons, but not in the bulk of *glr-1*-expressing neurons<sup>31</sup> (Fig. 4c). AIZ interneurons have not been reported to express endogenous *glr-1*, so AIB interneurons are probably the relevant site of *glr-1* rescue. Ablating the AIB neurons in the transgenic strain eliminated rescue, consistent with reconstitution of AIB function by the transgene (Fig. 4c). Ablation of AIZ interneurons in the transgenic strain had the same moderate effect as ablating AIZ in the wild type, indicating that AIZ



**Figure 2 | Both classes of AWC neurons respond to odour removal.**

**a–d**, G-CaMP calcium responses of AWC<sup>ON</sup> (**a**, **b**) or AWC<sup>OFF</sup> (**c**, **d**) after removal from different concentrations of isoamyl alcohol (IAA; **a**, **c**) or benzaldehyde (BZ; **b**, **d**). **e**, Removal of 2,3-pentanedione activates AWC<sup>OFF</sup>, but not AWC<sup>ON</sup>. **f**, AWC<sup>ON</sup> calcium responses in animals removed from isoamyl alcohol after different times of exposure. **g**, **h**, Secondary responses in

AWC<sup>ON</sup> after isoamyl alcohol was removed at 10 s, added back at 20 s (**g**) or 40 s (**h**), and then removed again 20 s later. Traces represent averages of 7–30 individual recordings (Supplementary Fig. 2). Bar graphs, average percentage change during the 10 s after odour removal (highlighted in yellow). Error bars and the shaded region around the curves represent s.e.m. The colour key for bar graphs is the same as that for the corresponding response curve.

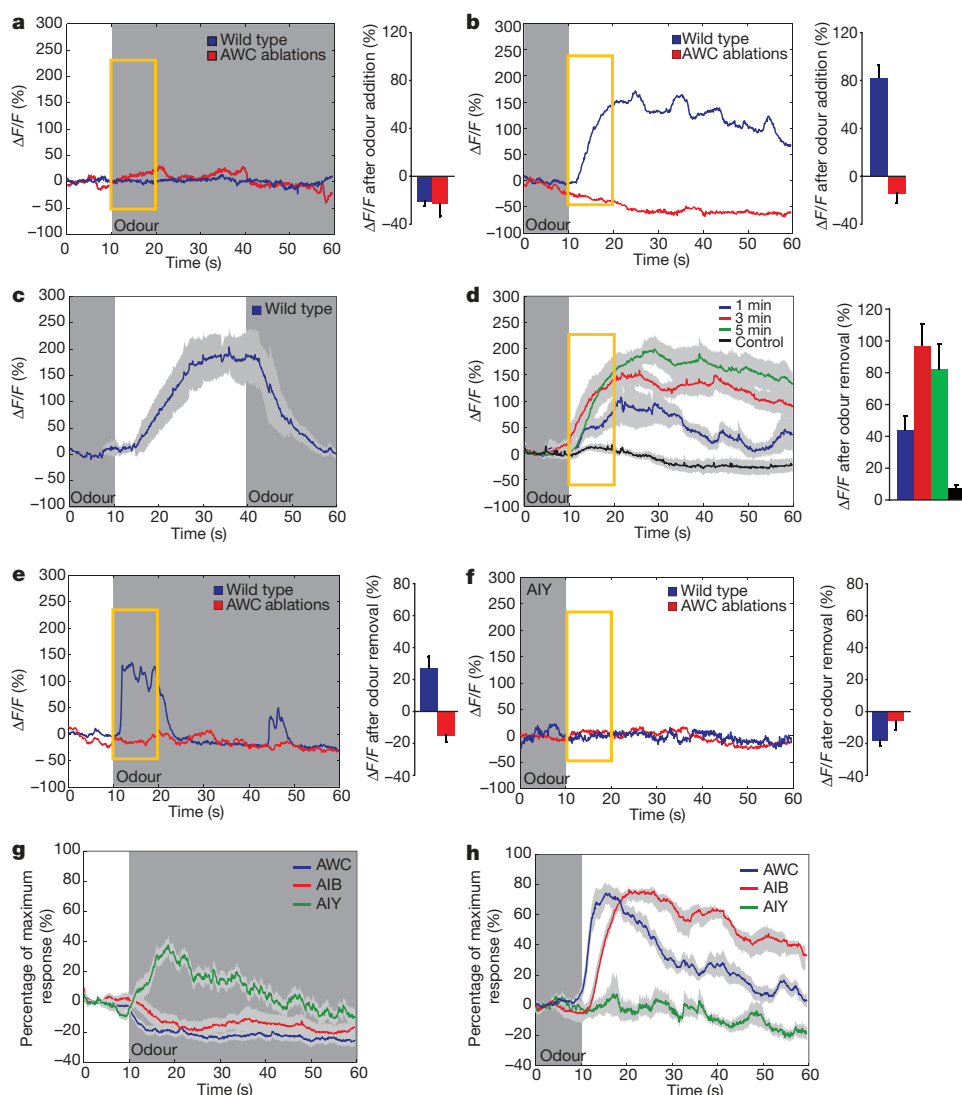
function was not disrupted by ectopic *glr-1* expression (Fig. 4c and Supplementary Fig. 6d). These results indicate that AWC releases glutamate to stimulate *glr-1* on AIB interneurons. The *glr-1(n2461)* mutant defect is less severe than the defect caused by killing AIB, suggesting roles for additional glutamate receptors in AIB interneurons or glutamate-independent activity of AIB interneurons.

AIY neurons do not express excitatory *glr* or *nmr* genes, but do express the glutamate-gated chloride channel GLC-3 (refs 32 and 33). *glc-3(ok321)* mutants had diminished local search behaviour that was rescued by expressing *glc-3* from an AIY-specific promoter (Fig. 4d). Killing AIY interneurons results in the opposite effect—an amplified and long-lasting local search behaviour<sup>14</sup> (Supplementary Fig. 6d). These results indicate that AWC inhibits AIY during local search behaviour via glutamate release onto GLC-3. *glr-1(n2461);glc-3(ok321)* double mutants were more defective than single mutants, consistent with parallel functions of AIB and AIY neurons (Fig. 4e).

To investigate this circuit model further, we used gain-of-function analysis in wild-type animals to test its predictions. In the vertebrate central nervous system, increased expression of synaptic AMPA-type

glutamate receptors results in synaptic strengthening during long-term potentiation<sup>34</sup>. Building from this premise, we asked whether overexpression of glutamate receptors would strengthen behavioural connections between AWC, AIY and AIB neurons. Increased expression of *glc-3* from an AIY promoter or *glr-1* from an AIB promoter resulted in animals with increased turning during local search (Fig. 4f), as did overexpressing both *AIY::glc-3* and *AIB::glr-1* (Fig. 4g). Killing the AWC sensory neurons abolished these turning behaviours, indicating that the extra turns were evoked by sensory signals (Fig. 4g). Control experiments showed that misexpression of the AIB receptor in AIY interneurons and vice versa disrupted local search behaviour (Fig. 4h).

Calcium imaging of interneurons in mutants provided further evidence for glutamate-mediated signalling between AWC, AIB and AIY neurons. Neuronal calcium signals and local search behaviours are measured on different timescales (1 min versus >5 min, respectively), but results from both approaches were analogous. AIB calcium responses to odour removal were eliminated in animals lacking the vesicular glutamate transporter EAT-4, but were partially



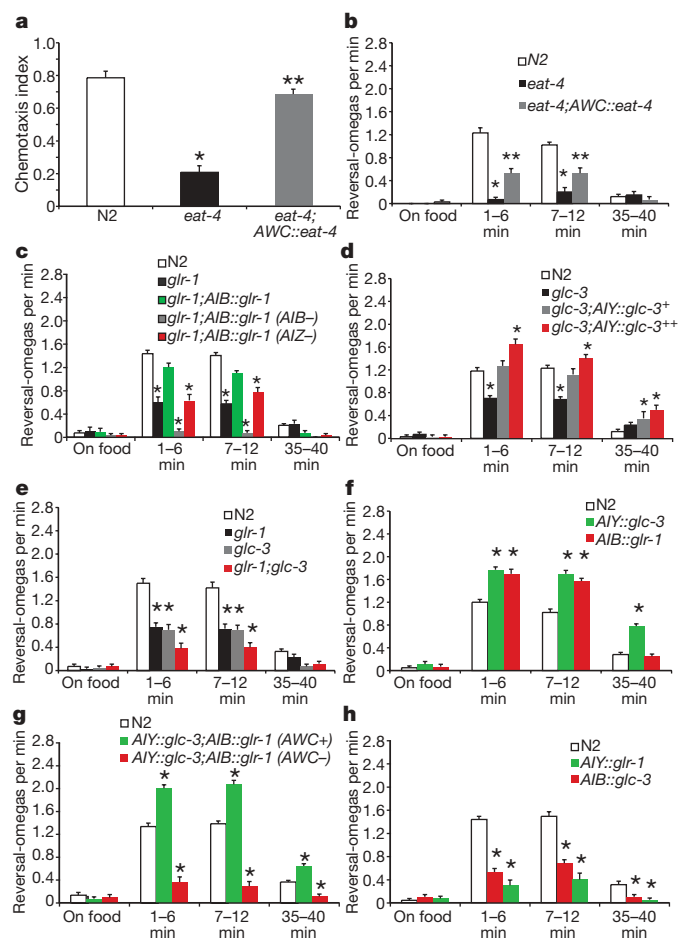
**Figure 3 | Calcium responses in AIB and AIY interneurons.**

**a, b**, Representative and average calcium responses of AIB interneurons in wild-type and AWC-ablated animals on odour addition (**a**) or removal (**b**). Bar graphs, average percentage change during the 10 s after odour addition or removal (highlighted in yellow;  $n = 22$  wild type, 7 AWC-ablated). **c**, Average AIB responses on odour removal (10 s) and re-addition (40 s). **d**, Average AIB responses to odour removal after different exposure times. **e, f**, Representative and average calcium responses of AIY neurites

from wild-type and AWC-ablated animals on odour addition (**e**) or removal (**f**) ( $n = 35$  wild type, 8 AWC-ablated). **g, h**, Averaged AWC<sup>ON</sup> ( $n = 30$ ), AIB ( $n = 59$ ) and AIY ( $n = 35$ ) responses, each trace normalized to its maximum response, on odour addition (**g**) and removal (**h**). Shading denotes the presence of odour. The shaded region around curves (**c, d, g and h**) indicates s.e.m. The colour key for bar graphs is the same as that for the corresponding response curve. Error bars indicate s.e.m.



restored by rescue of *eat-4* in AWC neurons (Fig. 5a, b). AIB responses were also eliminated in animals lacking the excitatory glutamate receptor GLR-1, and partially restored by rescue of *glr-1* in AIB neurons (Fig. 5c, d). The partial rescue may result from variable expression of transgenes, which is common in *C. elegans*, but it is likely that *eat-4* and *glr-1* also function in neurons other than AWC and AIB (for example, Supplementary Fig. 6b, c). The odour-induced increase in AIY calcium signals was eliminated in *glc-3(ok321)* mutant animals, supporting the identification of GLC-3 as an inhibitory postsynaptic receptor in AIY (Fig. 5e). However, in some animals (10 out of 22), an excitatory AIY calcium response was observed on odour removal (Fig. 5f and Supplementary Fig. 5e–h). These results indicate that AIY senses glutamate released from AWC neurons using the *glc-3* receptor and, in a *glc-3(ok321)* mutant background, may sense AWC signals through a second receptor with opposite activity.



**Figure 4 | AWC neurons signal through glutamate and glutamate receptors.** **a, b**, AWC-selective expression of *eat-4* rescues chemotaxis to isoamyl alcohol (**a**) and partially rescues local search turning behaviour after removal from food (**b**). Co-expression of *eat-4* in ASK neurons enhances rescue (Supplementary Fig. 6). N2, wild-type control strain. **c**, Local search turning behaviour of *glr-1(n2461)* mutants, rescue by AIB expression of *glr-1*, and effects of cell ablation (AIB<sup>-</sup> and AIZ<sup>-</sup>). **d**, Local search turning behaviour of *glc-3(ok321)* animals and rescue by AIY expression of *glc-3*. Some transgenic lines have an overexpression phenotype (++). **e**, *glr-1(n2461);glc-3(ok321)* double mutants are more defective than either single mutant. **f**, Enhanced turning in AIB::GLR-1 and AIY::GLC-3 animals overexpressing glutamate receptors in wild-type background. **g**, Enhanced turning in AIY::GLC-3;AIB::GLR-1 double-transgenic animals, and effects of AWC ablations (AWC<sup>-</sup>). **h**, Decreased turning in animals misexpressing *glr-1* in AIY or *glc-3* in AIB. Reversal-omega, paired reversal-omega turning sequences. Error bars indicate s.e.m. One asterisk, different from N2; two asterisks, different from *eat-4* ( $P < 0.05$ , Bonferroni *t*-test).

Together, these results indicate that, when AWC is active, it releases glutamate that acts on GLR-1 at the AWC::AIB synapse and GLC-3 at the AWC::AIY synapse, resulting in activation of AIB and inhibition of AIY. Because cell ablations indicate that AIB promotes turns and AIY inhibits turns during local search<sup>13,14</sup> (Supplementary Fig. 6d), both functions of AWC glutamate should promote turning.

### Odour removal evokes turning behaviour

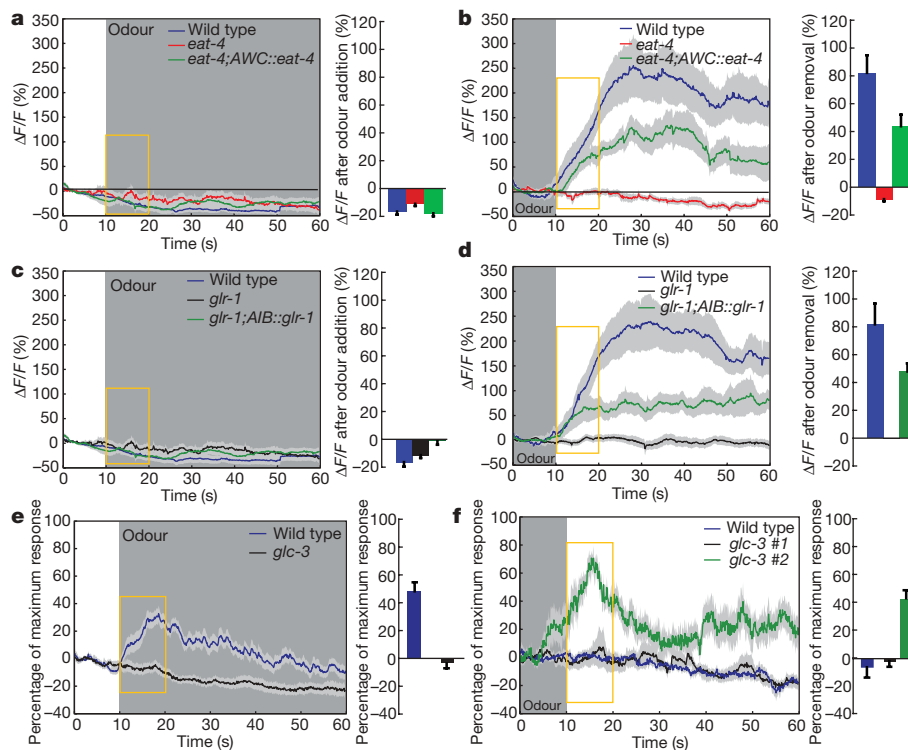
Food is a complex stimulus, and, although local search behaviour is highly dependent on AWC activity, it also includes contributions from other sensory neurons<sup>12–14</sup>. To tighten the correlation between odour responses observed by calcium imaging and turning behaviour elicited by AWC neurons, we asked whether removal of the AWC-sensed odour isoamyl alcohol stimulated turning. These experiments were inspired by the observation that chemotaxis to attractive salts in *C. elegans* is associated with regulated turning behaviour<sup>11,35</sup>. The turning behaviour of freely moving animals was monitored in a small chamber that was rapidly shifted between humidified air and humidified air equilibrated with the AWC-sensed odour isoamyl alcohol. Wild-type animals transiently increased turning when odour was removed, a response resembling local search behaviour, and transiently suppressed turning when odour was presented (Fig. 6a). These responses were diminished in *eat-4(ky5)* mutants with defective glutamate-mediated transmission, but were rescued by the expression of *eat-4* in AWC neurons (Fig. 6a). Enhanced turning after odour removal had a shorter duration than local search (2–5 min) but, like local search, it was impaired in *glr-1(n2461);glc-3(ok321)* glutamate receptor double mutants (Fig. 6b). These results suggest that *glr-1* and *glc-3* function in odour-regulated turning behaviour induced by AWC neurons.

The suppression of turning on odour addition was independent of *glr-1* and *glc-3*, yet dependent on *eat-4* in AWC neurons, indicating that additional glutamate receptors act downstream of AWC neurons. Similarly, chemotaxis in odour gradients requires *eat-4*, but was only slightly impaired in *glr-1(n2461);glc-3(ok321)* double mutants (data not shown). Many predicted *C. elegans* glutamate-receptor genes are uncharacterized, but the AMPA-like receptor genes *glr-2* and *glr-5* are expressed in AIB, and *glr-2* is expressed in AIA neurons, so these genes are candidates for additional components of the odour response<sup>36</sup>.

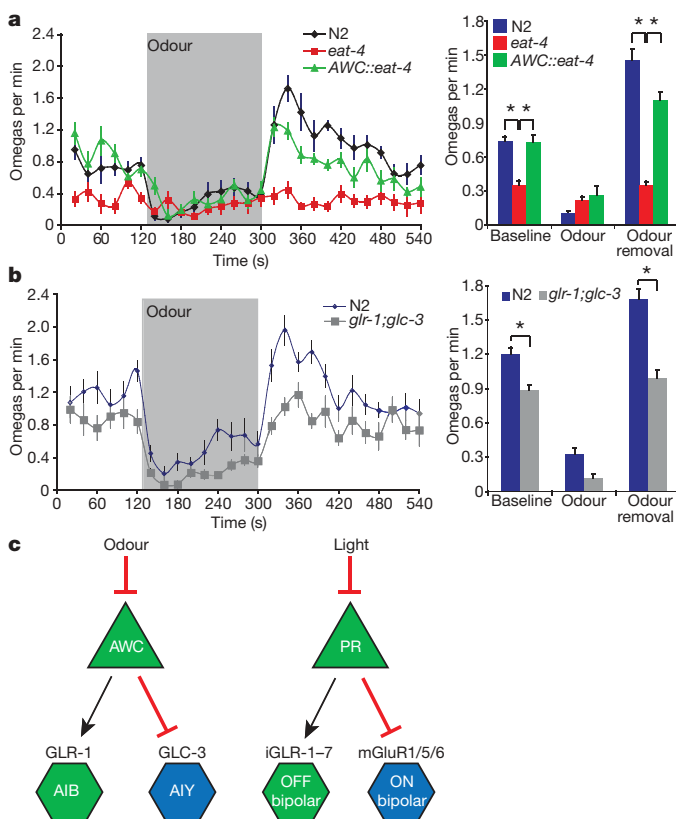
### Discussion

From these experiments, we infer that AWC olfactory neurons have basal activity in the absence of odour, are inhibited by odour, and are stimulated by odour removal. These graded properties resemble those of motor neurons in the nematode *Ascaris suum*, which are non-spiking, have tonic neurotransmitter release at rest, and respond to excitatory or inhibitory inputs with graded changes in membrane potential and transmitter release<sup>24</sup>. Thus, graded transmission may be a property of many nematode neurons.

The sensory responses of AWC neurons are similar to those of vertebrate rod and cone photoreceptors, which have tonic activity in the dark, are hyperpolarized by light, and are depolarized by the removal of light<sup>37</sup>. Like AWC neurons, photoreceptors are non-spiking and have graded glutamate release. Molecular analogies also link AWC neurons with vertebrate photoreceptors: their sensory transduction pathways rely on G-protein-coupled receptors, G<sub>i</sub>-like proteins, receptor-type guanylate cyclases and cyclic GMP-gated channels<sup>5,37</sup>, and their differentiation is controlled by Otx homeodomain proteins<sup>38,39</sup>. The synaptic connections between AWC, AIY and AIB neurons are also reminiscent of those between vertebrate photoreceptors and their targets, the ON and OFF bipolar cells<sup>40</sup> (Fig. 6c). In the retina, glutamate from photoreceptors is sensed by AMPA-type receptors on the OFF bipolar cell, a connection that is functionally and molecularly analogous to the AWC-to-AIB connection<sup>41</sup>. The vertebrate ON bipolar cell is inhibited by glutamate, as are AIY



**Figure 5 | AIB and AIY require different glutamate receptors.** **a, b,** Calcium responses of AIB interneurons in wild type, *eat-4(ky5)* mutants, and *eat-4(ky5)* mutants with an *AWC::eat-4* rescuing transgene on addition (**a**) or removal (**b**) of the odour isoamyl alcohol ( $n = 12$  for each). **c, d,** Calcium responses of AIB interneurons in wild type, *glr-1(n2461)* mutants, and *glr-1(n2461)* mutants with an *AIB::glr-1* rescuing transgene on addition (**c**) or removal (**d**) of isoamyl alcohol ( $n = 27$  for wild type,  $n = 25$  for *glr-1*,  $n = 12$  for *AIB::glr-1*). **e, f,** Calcium responses of AIY neurites in wild type and *glc-3(ok321)* mutants on addition (**e**) or removal (**f**) of odour ( $n = 35$  for wild type, 22 for *glc-3*). Twelve out of 22 *glc-3(ok321)* mutants did not respond to odour (*glc-3#1*); 10 out of 22 *glc-3(ok321)* mutants responded to odour removal, unlike wild type (*glc-3#2*) (Supplementary Fig. 5). Shading denotes the presence of odour. Bar graphs, average percentage change during the 10 s after odour addition or removal (highlighted in yellow). Error bars and shaded regions around curves represent s.e.m. The colour key for bar graphs is the same as that for the corresponding response curve.



neurons. In mammals, the ON bipolar cell is inhibited by a G-protein-coupled glutamate receptor and, in fish, by a glutamate-gated chloride channel that is functionally although not molecularly similar to *C. elegans* GLC-3 (ref. 42). The parallel ON and OFF streams enhance contrast sensitivity in vertebrate vision<sup>43</sup>; it is possible that the parallel AIY and AIB neurons have analogous functions in odour detection.

Like AWC sensory neurons, AIB interneurons respond to odour removal, but, unlike the transient AWC calcium signals, AIB calcium responses are sustained for several minutes. These longer-lasting responses indicate that AIB interneurons integrate the AWC response over time (Supplementary Fig. 7). Persistent neural activity is a common feature of neural circuits, and can arise from cell-intrinsic or network mechanisms<sup>23</sup>. The persistent activity of AIB may contribute to sustained turning behaviour after odour or food removal.

Individual turns after odour removal are stochastic and infrequent; they are not synchronized to odour removal and AIB activation. This mismatch suggests that the AIB could be a gating neuron

**Figure 6 | Odour-regulated turning behaviours.** **a, b,** Omega turns of freely moving animals recorded before, during and after a 3-min pulse of the odour isoamyl alcohol ( $t = 120$ – $300$  s). Odour addition suppressed turns and odour removal enhanced turns in wild type N2 (Supplementary Fig. 6e, f). **a,** *eat-4(ky5)* mutants and rescue by AWC expression of *eat-4*. **b,** *glr-1(n2461);glc-3(ok321)* glutamate receptor double mutants. Bar graphs, omega turns during 1 min before odour addition (baseline), after odour addition and after odour removal. Asterisks, values different from *eat-4* (**a**) or wild type (**b**) at  $P < 0.01$  (Bonferroni *t*-test). Error bars indicate s.e.m. **c,** Schematic comparing AWC, AIB and AIY neurons to vertebrate photoreceptors and bipolar neurons. iGluR, ionotropic glutamate receptor; mGluR, metabotropic glutamate receptor.



that facilitates turns, but does not trigger individual turning events<sup>44</sup>. The circuitry downstream of AIB interneurons may generate turns only when AIB is active and other unknown conditions are met. Possible sources for the turning signal are the secondary calcium transients in AWC neurons and the irregular, transient activity of AIY interneurons. Additional AWC target neurons such as AIA interneurons may also regulate turning; so far, the mechanism by which the AWC neuron suppresses turns on odour addition is unexplained, suggesting the existence of other target cells or receptors.

These experiments provide the beginnings of a circuit-level explanation for food-seeking behaviour in *C. elegans*. The regulation of turning by AWC neurons has features in common with the biased random walk, first described in bacterial chemotaxis and subsequently applied to *C. elegans* salt chemotaxis<sup>11,45,46</sup>. In this strategy, the animal senses temporal changes in attractant concentrations as it moves, resulting in long runs towards the attractant, short runs away from the attractant, and eventual accumulation at the point source. For AWC neurons, turning regulation can generate either chemotaxis or undirected local search; in both cases, turns are probabilistic and are not coupled tightly to stimuli. Our results and those of others suggest that the probabilistic turning circuit described here is both general and flexible. For example, temperature and salt stimuli affect *C. elegans* behaviour in a manner consistent with the biased random walk model<sup>35,46–48</sup>, and the temperature-sensing and salt-sensing neurons synapse onto AIB, AIY, AIA and/or AIZ interneurons<sup>3,25,49</sup>. A serotonin receptor essential for olfactory learning acts in AIY, AIB and AIZ interneurons, suggesting that the circuit is subject to plasticity<sup>50</sup>. Further studies of the circuit should shed light on the mechanism for stochastic behaviours, the generation of persistent behavioural responses to transient stimuli, and temporal integration in the nervous system.

## METHODS SUMMARY

Neuronal calcium responses were measured by detecting changes in fluorescence of G-CaMP, a non-ratiometric calcium indicator<sup>16</sup>; control experiments indicated that G-CaMP performed comparably to the ratiometric indicator cameleon in this system<sup>19</sup> (Supplementary Methods). Custom-designed microfluidic devices for trapping, stimulating and imaging calcium transients in *C. elegans* were fabricated from PDMS<sup>19</sup>. Fluid streams under laminar flow were used to deliver buffer or odour stimuli to the animal's nose, using flow by means of two alternative side-streams to control the delivery of stimuli without changing fluid pressure. Images were captured using a Coolsnap HQ camera and were analysed using Metamorph software.

Exploratory behaviours were scored essentially as described<sup>14</sup>. Animals were removed from food to a covered, food-free nematode growth medium (NGM) plate. Their behaviour was observed directly, and turning events were recorded using a Perl script. At least eight animals were scored blind for each condition.

Odour-induced turning responses were monitored in a custom-fabricated Plexiglass chamber over an agar surface, with an inlet, an outlet and a central imaging arena. Humidified air with or without odour was delivered to >20 freely moving animals using an electronic valve. Movies of these animals were analysed using automated tracking software, which is distributed at <http://wormsense.stanford.edu/tracker>. The software automatically identifies tens of animals, computes and tracks the centroid of each animal, and identifies turns on the basis of changes in angular velocity (see Supplementary Methods).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 14 August; accepted 24 September 2007.

- Milo, R. *et al.* Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
- Chalfie, M. *et al.* The neural circuit for touch sensitivity in *Caenorhabditis elegans*. *J. Neurosci.* **5**, 956–964 (1985).
- White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Phil. Trans. R. Soc. Lond. B* **314**, 1–340 (1986).
- de Bono, M. & Maricq, A. V. Neuronal substrates of complex behaviors in *C. elegans*. *Annu. Rev. Neurosci.* **28**, 451–501 (2005).
- Bargmann, C. I. In *Wormbook* (ed. The *C. elegans* Research Community) WormBook doi/10.1895/wormbook.1.123.1 (<http://www.wormbook.org>) (2006).

- Zheng, Y., Brockie, P. J., Mellem, J. E., Madsen, D. M. & Maricq, A. V. Neuronal control of locomotion in *C. elegans* is modified by a dominant mutation in the GLR-1 ionotropic glutamate receptor. *Neuron* **24**, 347–361 (1999).
- Tobin, D. *et al.* Combinatorial expression of TRPV channel proteins defines their sensory functions and subcellular localization in *C. elegans* neurons. *Neuron* **35**, 307–318 (2002).
- Goodman, M. B., Hall, D. H., Avery, L. & Lockery, S. R. Active currents regulate sensitivity and dynamic range in *C. elegans* neurons. *Neuron* **20**, 763–772 (1998).
- Kerr, R. *et al.* Optical imaging of calcium transients in neurons and pharyngeal muscle of *C. elegans*. *Neuron* **26**, 583–594 (2000).
- Mellem, J. E., Brockie, P. J., Zheng, Y., Madsen, D. M. & Maricq, A. V. Decoding of polymodal sensory stimuli by postsynaptic glutamate receptors in *C. elegans*. *Neuron* **36**, 933–944 (2002).
- Pierce-Shimomura, J. T., Morse, T. M. & Lockery, S. R. The fundamental role of pirouettes in *Caenorhabditis elegans* chemotaxis. *J. Neurosci.* **19**, 9557–9569 (1999).
- Hills, T., Brockie, P. J. & Maricq, A. V. Dopamine and glutamate control area-restricted search behavior in *Caenorhabditis elegans*. *J. Neurosci.* **24**, 1217–1225 (2004).
- Wakabayashi, T., Kitagawa, I. & Shingai, R. Neurons regulating the duration of forward locomotion in *Caenorhabditis elegans*. *Neurosci. Res.* **50**, 103–111 (2004).
- Gray, J. M., Hill, J. J. & Bargmann, C. I. A circuit for navigation in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **102**, 3184–3191 (2005).
- Zhao, B., Khare, P., Feldman, L. & Dent, J. A. Reversal frequency in *Caenorhabditis elegans* represents an integrated response to the state of the animal and its environment. *J. Neurosci.* **23**, 5319–5328 (2003).
- Nakai, J., Ohkura, M. & Imoto, K. A high signal-to-noise Ca<sup>2+</sup> probe composed of a single green fluorescent protein. *Nature Biotechnol.* **19**, 137–141 (2001).
- Pologruto, T. A., Yasuda, R. & Svoboda, K. Monitoring neural activity and [Ca<sup>2+</sup>] with genetically encoded Ca<sup>2+</sup> indicators. *J. Neurosci.* **24**, 9572–9579 (2004).
- Jospin, M., Jacquemond, V., Mariol, M. C., Segalat, L. & Allard, B. The L-type voltage-dependent Ca<sup>2+</sup> channel EGL-19 controls body wall muscle function in *Caenorhabditis elegans*. *J. Cell Biol.* **159**, 337–348 (2002).
- Chronis, N., Zimmer, M. & Bargmann, C. I. Microfluidics for *in vivo* imaging of neuronal and behavioral activity in *Caenorhabditis elegans*. *Nature Methods* **4**, 727–731 (2007).
- Bargmann, C. I., Hartwig, E. & Horvitz, H. R. Odorant-selective genes and neurons mediate olfaction in *C. elegans*. *Cell* **74**, 515–527 (1993).
- Colbert, H. A. & Bargmann, C. I. Environmental signals modulate olfactory acuity, discrimination, and memory in *Caenorhabditis elegans*. *Learn. Mem.* **4**, 179–191 (1997).
- Wes, P. D. & Bargmann, C. I. *C. elegans* odour discrimination requires asymmetric diversity in olfactory neurons. *Nature* **410**, 698–701 (2001).
- Major, G. & Tank, D. Persistent neural activity: prevalence and mechanisms. *Curr. Opin. Neurobiol.* **14**, 675–684 (2004).
- Davis, R. E. & Stretton, A. O. W. Signalling properties of *Ascaris* motor neurons: graded synaptic transmission and tonic transmitter release. *J. Neurosci.* **9**, 415–425 (1989).
- Clark, D. A., Biron, D., Sengupta, P. & Samuel, A. D. The AFD sensory neurons encode multiple functions underlying the mototactic behavior in *Caenorhabditis elegans*. *J. Neurosci.* **26**, 7444–7451 (2006).
- Lee, R. Y., Sawin, E. R., Chalfie, M., Horvitz, H. R. & Avery, L. EAT-4, a homolog of a mammalian sodium-dependent inorganic phosphate cotransporter, is necessary for glutamatergic neurotransmission in *Caenorhabditis elegans*. *J. Neurosci.* **19**, 159–167 (1999).
- Hart, A. C., Sims, S. & Kaplan, J. M. Synaptic code for sensory modalities revealed by *C. elegans* GLR-1 glutamate receptor. *Nature* **378**, 82–85 (1995).
- Maricq, A. V., Peckol, E., Driscoll, M. & Bargmann, C. I. Mechanosensory signalling in *C. elegans* mediated by the GLR-1 glutamate receptor. *Nature* **378**, 78–81 (1995).
- Cully, D. F. *et al.* Cloning of an avermectin-sensitive glutamate-gated chloride channel from *Caenorhabditis elegans*. *Nature* **371**, 707–711 (1994).
- Dillon, J., Hopper, N. A., Holden-Dye, L. & O'Connor, V. Molecular characterization of the metabotropic glutamate receptor family in *Caenorhabditis elegans*. *Biochem. Soc. Trans.* **34**, 942–948 (2006).
- Chou, J. H., Bargmann, C. I. & Sengupta, P. The *Caenorhabditis elegans odr-2* gene encodes a novel Ly-6-related protein required for olfaction. *Genetics* **157**, 211–224 (2001).
- Horoszok, L., Raymond, V., Sattelle, D. B. & Wolstenholme, A. J. GLC-3: a novel fipronil and BDNF-sensitive, but picrotoxin-insensitive, L-glutamate-gated chloride channel subunit from *Caenorhabditis elegans*. *Br. J. Pharmacol.* **132**, 1247–1254 (2001).
- Wenick, A. S. & Hobert, O. Genomic cis-regulatory architecture and trans-acting regulators of a single interneuron-specific gene battery in *C. elegans*. *Dev. Cell* **6**, 757–770 (2004).
- Malinow, R. & Malenka, R. C. AMPA receptor trafficking and synaptic plasticity. *Annu. Rev. Neurosci.* **25**, 103–126 (2002).
- Miller, A. C., Thiele, T. R., Faumont, S., Moravec, M. L. & Lockery, S. R. Step-response analysis of chemotaxis in *Caenorhabditis elegans*. *J. Neurosci.* **25**, 3369–3378 (2005).
- Brockie, P. J., Madsen, D. M., Zheng, Y., Mellem, J. & Maricq, A. V. Differential expression of glutamate receptor subunits in the nervous system of

- Caenorhabditis elegans* and their regulation by the homeodomain protein UNC-42. *J. Neurosci.* **21**, 1510–1522 (2001).
37. Zhang, X. & Cote, R. H. cGMP signaling in vertebrate retinal photoreceptor cells. *Front. Biosci.* **10**, 1191–1204 (2005).
  38. Furukawa, T., Morrow, E. M. & Cepko, C. L. *Crx*, a novel otx-like homeobox gene, shows photoreceptor-specific expression and regulates photoreceptor differentiation. *Cell* **91**, 531–541 (1997).
  39. Lanjuin, A., VanHoven, M. K., Bargmann, C. I., Thompson, J. K. & Sengupta, P. *Otx/otd* homeobox genes specify distinct sensory neuron identities in *C. elegans*. *Dev. Cell* **5**, 621–633 (2003).
  40. Yang, X.-L. Characterization of receptors for glutamate and GABA in retinal neurons. *Prog. Neurobiol.* **73**, 127–150 (2004).
  41. Wässle, H. Parallel processing in the mammalian retina. *Nature Rev. Neurosci.* **5**, 747–757 (2004).
  42. Grant, G. B. & Dowling, J. E. A glutamate-activated chloride current in cone-driven ON bipolar cells of the white Perch retina. *J. Neurosci.* **15**, 3852–3862 (1995).
  43. Schiller, P. H., Sandell, J. H. & Maunsell, J. H. Functions of the ON and OFF channels of the visual system. *Nature* **322**, 824–825 (1986).
  44. Taha, S. A. & Fields, H. L. Inhibitions of nucleus accumbens neurons encode a gating signal for reward-directed behavior. *J. Neurosci.* **26**, 217–222 (2006).
  45. Berg, H. C. & Brown, D. A. Chemotaxis in *Escherichia coli* analysed by three-dimensional tracking. *Nature* **239**, 500–504 (1972).
  46. Dusenbery, D. B. Responses of the nematode *Caenorhabditis elegans* to controlled chemical stimulation. *J. Comp. Physiol.* **136**, 327–331 (1980).
  47. Ryu, W. S. & Samuel, A. D. Thermotaxis in *Caenorhabditis elegans* analyzed by measuring responses to defined thermal stimuli. *J. Neurosci.* **22**, 5727–5733 (2002).
  48. Zariwala, H. A., Miller, A. C., Faumont, S. & Lockery, S. R. Step response analysis of thermotaxis in *Caenorhabditis elegans*. *J. Neurosci.* **23**, 4369–4377 (2003).
  49. Mori, I. & Ohshima, Y. Neural regulation of thermotaxis in *Caenorhabditis elegans*. *Nature* **376**, 344–348 (1995).
  50. Zhang, Y., Lu, H. & Bargmann, C. I. Pathogenic bacteria induce aversive olfactory learning in *Caenorhabditis elegans*. *Nature* **438**, 179–184 (2005).
- Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).
- Acknowledgements** We thank the *C. elegans* Knockout Consortium and the *Caenorhabditis* Genetic Center (CGC) for strains, A. Wolstenholme for the *glc-3* cDNA, P. Sengupta for the *srsx-3* promoter, and M. Meister, M. Zimmer, B. Snyder, G. Lee, D. Albrecht, M. Hilliard and other Bargmann laboratory members for critical help, insights and advice. S.H.C. was supported by the Damon Runyon Cancer Research Foundation and C.I.B. is an Investigator of the Howard Hughes Medical Institute. This work was supported by the Howard Hughes Medical Institute (C.I.B.) and the Klingenstein Fund for Neuroscience (M.B.G.).
- Author Contributions** S.H.C. designed and performed experiments, analysed data and wrote the paper; N.C., M.T. and J.M.G. designed and performed experiments; D.R. and M.B.G. developed analytical tools; and C.I.B. designed experiments, analysed data and wrote the paper.
- Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to C.I.B. ([cori@rockefeller.edu](mailto:cori@rockefeller.edu)).



## METHODS

**Calcium imaging.** The microfluidic device was fabricated using standard micro-machining procedures<sup>19,51</sup>. The design was drawn in Autocad and the chrome mask was generated by a mask-making service (Microlab, University of California, Berkeley). A master mould was created by spin casting (~2,800 r.p.m.) and photolithographically patterning a 28- $\mu$ m-thick layer of SU-8-2025 photoresist on a bare silicon wafer. A PDMS prepolymer mixture (Sylgard 184, 10:1) was cast over the mould and cured on a hot plate for 2 h at 70 °C. PDMS devices were peeled off the mould, treated with air plasma (30 W for 30 s) and irreversibly bonded to a glass coverslip. A sharpened needle (0.16 inch inner diameter  $\times$  0.25 inch outer diameter) was used to create inlets and outlets for the PDMS chip. Syringes with buffers were connected to the microfluidic device using polyethylene tubing (0.58 mm inner diameter  $\times$  0.95 mm outer diameter). Constant suction applied to the outlet of the chip generated laminar flow profiles for the four different streams (two dye (fluorescein) streams, one buffer stream and one stimulus stream). Individual worms were sucked into a buffer-filled polyethylene tube with the aid of a syringe and injected into the entrance of the chip by manually increasing the pressure by means of the syringe. A three-way valve (The Lee Company, 778360) regulated the two dye streams allowing only one dye stream into the chip at any given time. In the OFF condition (Fig. 1c, channels 1–3 open), the dye stream (flow 1) pushed the stimulus away from the nose, immersing the nose in the buffer stream (flow 2). In the ON condition (Fig. 1d, channels 2–4 open), dye stream 4 pushed stimulus stream 3 (odour or bacterial-conditioned medium of  $A_{600}$  1.0 in buffer) towards the nose of the trapped animal. Indirect control of fluid streams was necessary because AWC neurons respond strongly to small changes in fluid pressure (data not shown). G-CaMP imaging was performed on a Zeiss Axioscope upright microscope using a Coolsnap HQ photometrics camera. A detailed description of image analysis is provided in the Supplementary Methods.

**Odour flow assays.** A custom-fabricated Plexiglass device with an inlet, an outlet and a central arena (30 mm  $\times$  44 mm wide and 0.3 mm high; air volume 3.6 ml) was placed on a food-free NGM assay plate to create a chamber for imaging movement. Animals were picked on the NGM plate and corralled in the arena with a filter paper soaked in the repellent copper chloride (20 mM CuCl<sub>2</sub>), which kept them in view for 30 min. The inlet was controlled by two three-way valves (The Lee Company, 778360) allowing air equilibrated in water or air equilibrated in isoamyl alcohol diluted in water to flow through the chamber at 200 ml min<sup>-1</sup>. The saturated vapour concentration of isoamyl alcohol is  $1.2 \times 10^{-4}$  M at 20 °C, corresponding to a  $10^{-5}$  dilution in liquid phase. Stimulus-equilibrated air was further diluted 1:10 in air before being presented to the behaviour chamber. A Zeiss dissecting microscope and a Macrofire SDK camera captured the behaviour of all animals. Images were analysed using an automatic tracking system (see Supplementary Methods). Results were averaged in 20-s bins.

**Quantitative analysis of exploratory behaviour.** Animals were scored for exploratory behaviour on food, immediately after removal from food, and at long times off food. Adult animals were first observed and scored for 5 min on a lawn of OP50 *Escherichia coli* bacteria on an NGM plate and then transferred to a food-free NGM plate, where they were observed and scored during 15 of the next 40 min. The assay was performed as described<sup>14</sup> except that the food-free assay plate was covered by a lid during the assay. All turns and reversals were scored by eye, by an investigator blind to the genotype and ablation status of the animal. Reversals that had three or more head swings were identified as long reversals; these are essentially absent in the presence of food (Supplementary Fig. 6a, b)<sup>14</sup>. Turns in which the head nearly touched the tail or turns in which a single head swing led to a reorientation of more than 135° were identified as omega turns (Supplementary Fig. 6a)<sup>14</sup>. Results in Fig. 4 are reported as RevOmega values, which represent coupled large reversal-omega pairs, but qualitatively similar results were obtained when either large reversals or omega turns were scored individually (Supplementary Table 1). Reversals coupled to omega turns were chosen for display because of their consistency in wild-type animals across various experiments. Data were analysed using Perl scripts to calculate reversal and omega frequencies.

51. Younan, X. & Whitesides, G. Soft lithography. *Ann. Rev. Mater. Sci.* **28**, 153–184 (1998).

# Circular polarimetry reveals helical magnetic fields in the young stellar object HH 135–136

Antonio Chrysostomou<sup>1</sup>, Philip W. Lucas<sup>1</sup> & James H. Hough<sup>1</sup>

Magnetic fields are believed to have a vital role in regulating and shaping the flow of material onto and away from protostars during their initial mass accretion phase. It is becoming increasingly accepted<sup>1</sup> that bipolar outflows are generated and collimated as material is driven along magnetic field lines and centrifugally accelerated off a rotating accretion disk. However, the precise role of the magnetic field is poorly understood and evidence for its shape and structure has not been forthcoming. Here we report imaging circular polarimetry in the near-infrared and Monte Carlo modelling showing that the magnetic field along the bipolar outflow of the HH 135–136 young stellar object is helical. The field retains this shape for large distances along the outflow, so the field structure can also provide the necessary magnetic pressure for collimation of the outflow. This result lends further weight to the hypothesis—central to any theory of star formation—that the outflow is an important instrument for the removal of high-angular-momentum material from the accretion disk, thereby allowing the central protostar to increase its mass.

Herbig–Haro (HH) objects were discovered as faint, nebulous emission on optical plates in the 1950s by G. H. Herbig<sup>2</sup> and G. Haro<sup>3</sup>. They possess characteristic optical spectra<sup>4</sup> and are believed to be the result of a bipolar jet or outflow driven from a young stellar object (YSO) interacting with its parent molecular cloud<sup>1</sup>. In fact, the astrophysical jet is considered to be of major importance given its manifestation in several different environments from brown dwarfs<sup>5</sup> and binary systems<sup>1</sup> through to black holes and active galactic nuclei<sup>5</sup>. Theoretical models that seek to explain the jet phenomenon now consistently appeal to the magnetohydrodynamic (MHD) interaction of the magnetic field with the accreting gas in some form or another: X-wind models<sup>6</sup> generate an outflow very close to the star (of the order of a few stellar radii) in the magnetopause between the accretion disk and the protostar; in contrast, disk-wind models<sup>6,7</sup> generate an outflow over a relatively large range of radii of the disk surface. In each case, material is lifted and centrifugally accelerated along magnetic field lines (rather like beads on a wire). Although present observational techniques struggle to resolve and probe into the innermost regions where the jet is launched, these theoretical models are able to explain some key observational properties, such as the correlation of accretion with ejection rates, and the fractional rates of mass loss<sup>7</sup>. Observations are beginning to favour disk-wind models<sup>8,9</sup> and show that the outflow remains collimated over large distances, although the debate is far from resolved.

Mechanical collimation is possible if the ram pressure in the flow (acting normal to the cavity surface) is equal to or less than the pressure in the medium that the outflow is pushing through. Although such collimation is plausible for young sources that remain embedded within the natal molecular cloud, it does not explain the collimation seen in jets from optically revealed YSOs<sup>10</sup>—the so-called T Tauri stars—which have had their circumstellar environment

cleared away. To reconcile this inconsistency, theories invoke the presence and action of magnetic fields by which an initially collapsing cloud forms a rotating accretion disk in the plane perpendicular to the magnetic field lines, which in turn twists the field lines into a helical field (with opposite helicities above and below the disk). This can provide the necessary magnetic pressure in the environment to retain the observed collimation to large distances by means of ‘hoop’ stresses<sup>11</sup>. Although evidence for strong ( $>1$ -mG) magnetic fields at large distances from YSOs is available<sup>12</sup>, there have been few such observations, and evidence for the morphology of this field is severely lacking.

The efficacy of polarimetry in constraining various physical parameters concerning the scattering geometry and media in the environments of young stellar objects has been shown in numerous papers<sup>13,14</sup>. Grains are in general non-spherical and aligned by the ambient magnetic field, and selective absorption by the long axis of the grains polarizes the radiation (dichroism). Linear polarization (generated by dust scattering and/or dichroism) has an almost ubiquitous presence in star formation regions, whereas circular polarization is far less common and requires more specialized conditions for its generation, for example multiple scattering, scattering off aligned non-spherical grains (dichroic scattering) or dichroic extinction of linearly polarized light.

In Fig. 1 we show near-infrared imaging circular polarimetry data for the HH 135–136 outflow. Situated at a distance of 2.7 kpc in the Carina nebula<sup>15</sup> these are two of the most distant HH objects known, powered by an intermediate-mass Herbig Ae–Be star, IRAS 11101–5829 (ref. 16). The data show that the two lobes of the bipolar outflow are strongly circularly polarized. For the southern lobe, the bulk of the radiation is positively circularly polarized (that is, right-handed in the direction of propagation), whereas negative circularly polarized light dominates the northern lobe. In each case, flips in the handedness of circular polarization are seen towards the limbs of the outflow lobes (most apparent in the  $K_n$  and H band images). This pattern does not conform to the classical alternating symmetry seen in other objects and models<sup>13,14,17,18</sup>, in which the flip occurs along the axis of the flow and both senses of polarization are equally prominent. The degrees of circular polarization are relatively high. In the  $K_n$  band the circular polarization is as much as about 8%, becoming about 3% in the H band and about 2.5% in the J band. Values as high as ~15–20% in the  $K_n$  band have been reported towards the OMC-1 nebula in Orion<sup>13</sup> and in NGC6334V (ref. 19), both regions of high-mass star formation. The values we find here are significantly higher than those seen towards low-mass YSOs, for which absolute values  $<1.5\%$  are typically seen<sup>14,20</sup>. This places HH 135–136 in an intermediate position between those objects with high values of circular polarization, which are so far associated with high-mass YSOs, and those with low values, which are associated with low-mass YSOs. If one associates increasing magnetic field

<sup>1</sup>Centre for Astrophysics Research, Science and Technology Research Institute, University of Hertfordshire, Hatfield, Herts. AL10 9AB, UK.



strength with the formation of higher-mass stars, this apparent correlation can be understood simply, because in the presence of a strong magnetic field the grains will align more efficiently and will more readily polarize radiation circularly.

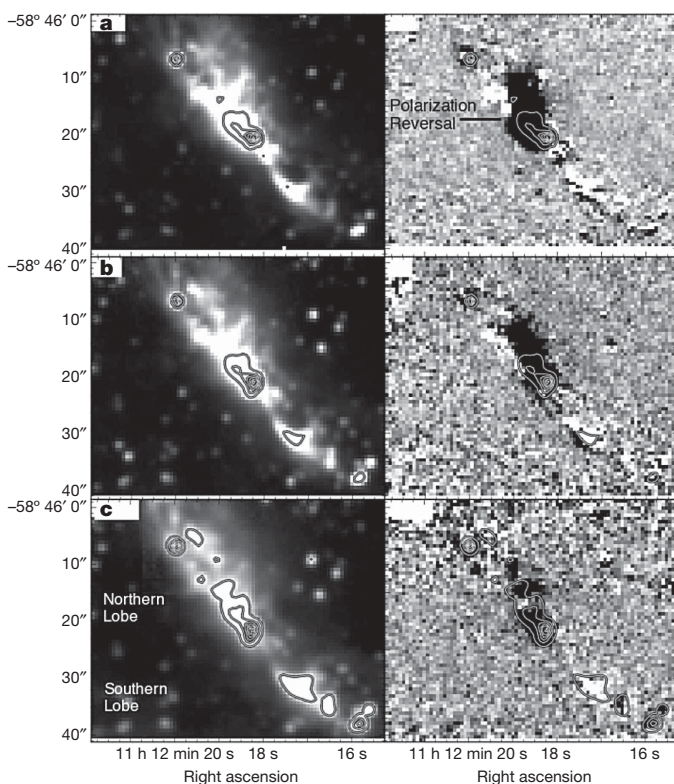
It seems most likely that the only way to produce the pattern observed, while still retaining the observed centrosymmetric linear polarization pattern produced by scattering of light from the central protostar<sup>16</sup>, is through the manipulation of the magnetic field structure. The classic symmetry of an alternating quadrupolar circular polarization pattern can arise from multiple scattering by spherical or non-aligned grains, but this produces only  $\sim 1\%$  polarization. If there are aligned non-spherical grains, the same pattern can be produced by dichroic scattering or through dichroic extinction if the field is oriented parallel to the outflow axis<sup>17,21,22</sup>, and the fractional polarization is higher.

Output from a successful model (see Supplementary Information) with a pinched and twisted field structure is shown in Fig. 2. We find that dichroic extinction (birefringence) is the principal cause of the observed 8% polarization. As well as reproducing the linear polarimetry<sup>16</sup>, the model reproduces key features of the circular

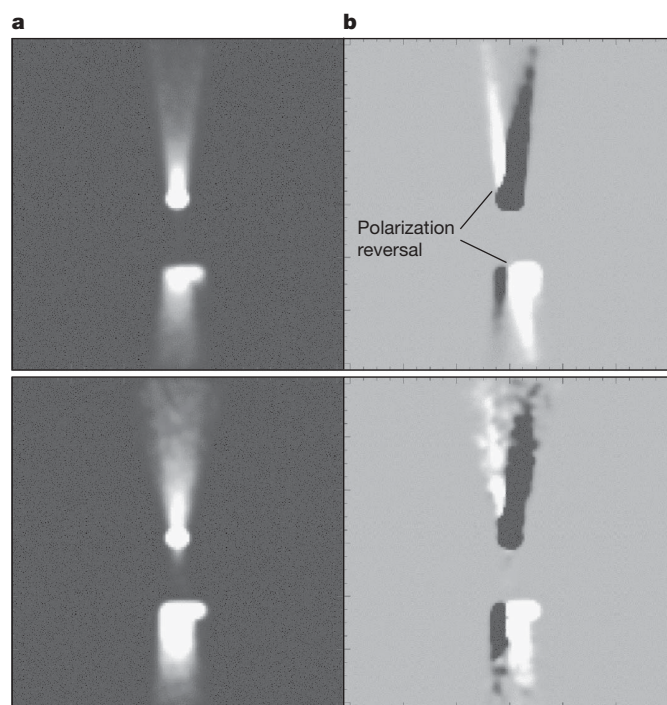
polarimetry: first, the opposite sense of polarization in the two outflow lobes, which requires that the helicity reverses in the disk plane (that is, a bihelical structure, which is naturally caused by a rotating disk); second, the flip in polarization at the eastern edge of the outflow in both lobes, at a certain distance along the flow, which is evidence that the pitch angle of the field to the disk plane increases with increasing distance (that is, the field becomes more axial); and last, the wavelength dependence of the polarization. This was contrary to the usual wavelength dependence of birefringence or dichroic scattering<sup>22,23</sup>, but it was found to be caused by the fairly high optical depth of the reflection nebosity, which was determined from the near-infrared colours ( $A_V \approx 13$ )<sup>16</sup>. At the shorter wavelengths the polarization is reduced by increased multiple scattering, because the optical depth and the grain albedo are higher at shorter wavelengths.

This optical depth implies a mass for the outflow of  $\sim 1M_\odot$ , but given the high  $H_2$  luminosity<sup>24</sup> and mass outflow rates<sup>25</sup> of objects typical of that driving HH 135–136 ( $\sim 10^{-3}$  to  $10^{-4} M_\odot \text{ yr}^{-1}$ ) this is not unfeasible. Optical and near-infrared images<sup>15,24</sup> show emission-line structures coincident with the reflection nebosity, indicating that we are tracing the material, and hence the magnetic field, in an outflow rather than an extended molecular envelope.

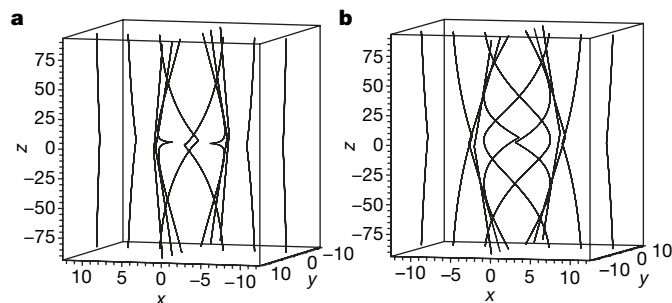
The field used is shown in Fig. 3. The model is not very sensitive to the radial component of the field (relative to the outflow axis) and



**Figure 1 | Imaging circular polarimetry of the HH 135–136 outflow system.** Polarimetry in the  $K_n$  (a), H (b) and J (c) near-infrared atmospheric passbands was obtained at the Anglo-Australian Telescope on the nights of 22–24 May 1997, using standard instrument and data reduction techniques<sup>14</sup>. The facility near-infrared imager-spectrometer (IRIS) was used at the  $f/15$  telescope focus. The left panels show the near-infrared intensity in the  $K_n$  ( $2.13 \mu\text{m}$  wavelength), H ( $1.6 \mu\text{m}$  wavelength) and J ( $1.25 \mu\text{m}$  wavelength) bands. Coordinates are given for epoch J (2000); north is to the top and east to the left. The right panels show the circular polarization (Stokes V) measured in these bands. Intensity contours were arbitrarily chosen for each waveband, to accentuate the brighter emission features, and are plotted on the intensity and circular polarization images. In each case, black is negatively circularly polarized and white is positive. By convention, positive polarization means that the electric vector is seen to rotate anticlockwise (right-handed) along the propagation direction in a fixed plane by an observer looking at the source. Each lobe is dominated by a single handedness of polarization and of opposite sign to the other lobe, although flips in the handedness of polarization are seen towards the edges of the lobes.



**Figure 2 | Results of three-dimensional Monte Carlo light-scattering model<sup>21</sup>.** a, Intensity image. b, Circular polarimetry image. The upper images are for the K band ( $2.2 \mu\text{m}$  wavelength) and the lower images are for the J band ( $1.25 \mu\text{m}$  wavelength). The light emanates from a central protostar and is scattered by dust in the bipolar outflow. The inner 50% of the outflow cross-section (by radius) is optically thin, so that the full length of the outflow is illuminated. The outer 50% is denser and is responsible for producing the circular polarization through dichroic extinction. The protostar itself is entirely obscured from view by an optically thick accretion disk (500 AU radius) and a circumstellar envelope (1,000 AU radius), which are viewed edge-on and are therefore also not seen. The bright peak near the centre of the model intensity images is due to scattering from the outflow in the upper lobe. The large vertical gap between the two outflow lobes is due to a much lower density of matter in the inner part of the lower lobe. The model outflow has a non-axisymmetric structure in the lower lobe, for a better reproduction of the data. The structure of the model is described fully in Supplementary Information.



**Figure 3 | Helical magnetic field structures.** **a**, The pinched and twisted magnetic field structure used in the model shown in Fig. 2. The toroidal component,  $B_\phi$ , slowly decays with distance from the source. This ensures that at large distances from the YSO, the magnetic field can remain contiguous with what is assumed to be a generally axial, large-scale, magnetic field running through the molecular cloud. **b**, A field with no pinch in the disk plane (that is,  $B_r = 0$ ) but with a slightly stronger twist. The two structures produce a similar circular polarization. The field is drawn to scale with the data;  $z = 100$  represents 27,000 AU, and  $z = 0$  represents the disk plane. The sign of  $B_z$  is the same in both hemispheres (that is, dipole polarity). The sense of rotation of the disk is such that its rotation axis is parallel to north. The equations describing the field structure are given in Supplementary Information.

so we cannot confirm whether the field lines are pinched in the disk plane, although the twist does seem to be strongest there. Observations at higher spatial resolution, to measure the flips in polarization at the eastern edge of the flow more precisely, could in principle measure the field pitch angle as a function of distance along the flow, although the imperfect axisymmetry of the outflow would introduce some uncertainty.

We suggest that a helical field is a plausible model for the magnetic field structure about HH 135–136 (and other YSO outflows) and possesses an appropriate configuration to provide the necessary transverse pressure to retain the collimation of outflows, confirming an earlier suggestion based on linear polarimetry<sup>26</sup>. We also suggest that circular polarimetry is a new and powerful technique for probing the magnetic field structure at these small scales (arcseconds on the sky) and calls for further observations to be made towards other YSOs. Finally, given that the combination of a rotating disk and a helical magnetic field structure would naturally launch material into the outflow with a significant toroidal component, angular momentum will be carried away from the central accreting system, thus preventing centrifugal forces from stopping the collapse and allowing material possessing a low angular momentum to be accreted onto the protostar.

Received 8 June; accepted 4 September 2007.

1. Ray, T. P., Dougados, C., Bacciotti, F., Eisloffel, J. & Chrysostomou, A. in *Protostars & Planets V* (eds Reipurth, B., Jewitt, D. & Keil, K.) 231–244 (Univ. of Arizona Press, Tucson, 2007).
2. Herbig, G. H. The spectrum of the nebulosity surrounding T Tauri. *Astrophys. J.* **111**, 11–14 (1950).
3. Haro, G. Herbig's nebulous objects near NGC 1999. *Astrophys. J.* **115**, 572 (1952).
4. Hartigan, P., Morse, J. A., Tumlinson, J., Raymond, J. & Heathcote, S. Hubble Space Telescope faint object spectrograph optical and ultraviolet spectroscopy of the bow shock HH 47A. *Astrophys. J.* **512**, 901–915 (1999).

5. Reipurth, B. & Bertout, C. (eds) *Herbig–Haro Flows and the Birth of Low-mass Stars*. IAU Symposium No. 182 (Kluwer, Dordrecht, 1997).
6. Shu, F. H., Najita, J. R., Shang, H. & Li, Z.-Y. in *Protostars & Planets IV* (eds Mannings, V., Boss, A. P. & Russell, S.) 789–813 (Univ. of Arizona Press, Tucson, 2000).
7. Ferreira, J. Magnetically driven jets from Keplerian accretion disks. *Astron. Astrophys.* **319**, 340–359 (1997).
8. Bacciotti, F. et al. Hubble Space Telescope STIS spectroscopy of the optical outflow from DG Tauri: structure and kinematics on subarcsecond scales. *Astrophys. J.* **537**, L49–L52 (2000).
9. Takami, M. et al. Detection of a warm molecular wind in DG Tauri. *Astron. Astrophys.* **416**, 213–219 (2004).
10. Dougados, C., Cabrit, S., Lavalley, C. & Ménard, F. T Tauri stars microjets resolved by adaptive optics. *Astron. Astrophys.* **357**, L61–L64 (2000).
11. Pudritz, R. E., Ouyed, R., Fendt, C. & Brandenburg, A. in *Protostars & Planets V* (eds Reipurth, B., Jewitt, D. & Keil, K.) 277–294 (Univ. of Arizona Press, Tucson, 2007).
12. Chrysostomou, A., Hough, J. H., Burton, M. G. & Tamura, M. Twisting magnetic fields in the core region of OMC-1. *Mon. Not. R. Astron. Soc.* **268**, 325–334 (1994).
13. Chrysostomou, A. et al. Polarimetry of young stellar objects—III. Circular polarimetry of OMC-1. *Mon. Not. R. Astron. Soc.* **312**, 103–115 (2000).
14. Chrysostomou, A. et al. Polarimetry of young stellar objects—II. Circular polarization of GSS 30. *Mon. Not. R. Astron. Soc.* **285**, 750–758 (1997).
15. Ogura, K. & Walsh, J. R. Spectacular Herbig–Haro objects in the Carina nebula. *Astrophys. J.* **400**, 248–259 (1992).
16. Tamura, M. et al. Identification of the exciting source of Herbig–Haro objects 135 and 136. *Mon. Not. R. Astron. Soc.* **287**, 894–898 (1997).
17. Whitney, B. & Wolff, M. J. Scattering and absorption by aligned grains in circumstellar environments. *Astrophys. J.* **574**, 205–231 (2002).
18. Lucas, P. W. Computation of light scattering in young stellar objects. *J. Quant. Spectrosc. Radiat. Transf.* **79**, 921–937 (2003).
19. Ménard, F., Chrysostomou, A., Gledhill, T. M., Hough, J. H. & Bailey, J. A. in *Bioastronomy 99: A New Era in the Search for Life*. ASP Conference Series, Vol. 213 (eds Lemarchand, G. & Meech, K.) 355–358 (Astronomical Society of the Pacific, San Francisco, 2000).
20. Gledhill, T. M., Chrysostomou, A. & Hough, J. H. Linear and circular imaging polarimetry of the Chamaeleon infrared nebula. *Mon. Not. R. Astron. Soc.* **282**, 1418–1436 (1996).
21. Lucas, P. W. et al. High-resolution imaging polarimetry of HL Tau and magnetic field structure. *Mon. Not. R. Astron. Soc.* **352**, 1347–1364 (2004).
22. Lucas, P. W. et al. UV circular polarization in star formation regions: the origin of homochirality? *Orig. Life Evol. Biosphs* **35**, 29–60 (2005).
23. Buschermöhle, M. et al. An extended search for circularly polarized infrared radiation from the OMC-1 region of Orion. *Astrophys. J.* **624**, 821–826 (2005).
24. Gredel, R. HH135/136—a luminous  $H_2$  outflow towards a high-mass protostar. *Astron. Astrophys.* **457**, 157–166 (2006).
25. Arce, H. G. et al. in *Protostars & Planets V* (eds Reipurth, B., Jewitt, D. & Keil, K.) 245–260 (Univ. of Arizona Press, Tucson, 2007).
26. Rodrigues, C. V., Hickel, G. R., Cerqueira, A. H. & Targon, C. G. Optical polarimetry of HH135/136. *Astron. J.* **133**, 2334–2342 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M. Tamura for discussions on HH135 and for providing us with his linear polarimetry data to aid our analyses. This work was supported by a grant from the UK Science and Technologies Facilities Council (to P.W.L.).

**Author Contributions** All authors contributed equally to this work. A.C. and J.H.H. conducted the observations at the telescope. A.C. reduced the data, and P.W.L. performed the Monte Carlo modelling. A.C. wrote the main paper, and P.W.L. wrote the Supplementary Information. All authors discussed the results and implications and commented on the manuscript at all stages.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to A.C. ([a.chrysostomou@herts.ac.uk](mailto:a.chrysostomou@herts.ac.uk)) or P.W.L. ([p.w.lucas@herts.ac.uk](mailto:p.w.lucas@herts.ac.uk)).

## LETTERS

## The rotating wind of the quasar PG 1700+518

S. Young<sup>1,2</sup>, D. J. Axon<sup>1,2</sup>, A. Robinson<sup>1,2</sup>, J. H. Hough<sup>2</sup> & J. E. Smith<sup>2</sup>

It is now widely accepted that most galaxies undergo an active phase, during which a central super-massive black hole generates vast radiant luminosities through the gravitational accretion of gas<sup>1,2</sup>. Winds launched from a rotating accretion disk surrounding the black hole are thought to play a critical role, allowing the disk to shed angular momentum that would otherwise inhibit accretion<sup>3,4</sup>. Such winds are capable of depositing large amounts of mechanical energy in the host galaxy and its environs, profoundly affecting its formation and evolution<sup>5–7</sup>, and perhaps regulating the formation of large-scale cosmological structures in the early Universe<sup>8,9</sup>. Although there are good theoretical grounds for believing that outflows from active galactic nuclei originate as disk winds<sup>10</sup>, observational verification has proven elusive. Here we show that structures observed in polarized light across the broad H $\alpha$  emission line in the quasar PG 1700+518 originate close to the accretion disk in an electron scattering wind. The wind has large rotational motions ( $\sim 4,000 \text{ km s}^{-1}$ ), providing direct observational evidence that outflows from active galactic nuclei are launched from the disks. Moreover, the wind rises nearly vertically from the disk, favouring launch mechanisms that impart an initial acceleration perpendicular to the disk plane.

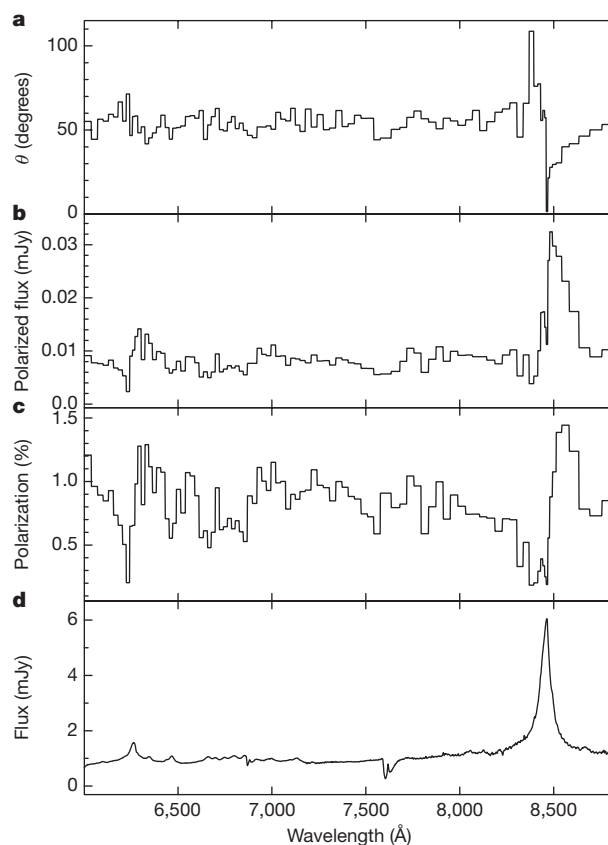
The blueshifted broad absorption lines (BALs) observed in the spectra of quasars provide some of the most compelling evidence for powerful outflows from active galactic nuclei (AGN)<sup>11</sup>. Although only about 15% of quasars exhibit BALs<sup>12</sup>, it is generally believed that the high velocity winds in which these features are formed are ubiquitous<sup>13</sup>, the BALs being observed only when our line of sight happens to intercept the wind, which occupies a relatively small solid angle.

Our optical spectropolarimetry of the low-redshift ( $z = 0.29$ ) BAL quasar PG 1700+518 (ref. 14) shows significant variations in both the degree ( $p(\lambda)$ ) and position angle ( $\theta(\lambda)$ ) of polarization across both the broad Balmer H $\alpha$  and H $\beta$  emission lines (Fig. 1). In polarized flux ( $p(\lambda) \times F(\lambda)$ ), the broad Balmer line profiles are significantly redshifted relative to their total flux counterparts. Relative to its value in the continuum,  $\theta(\lambda)$  deviates by  $30\text{--}40^\circ$ , rotating in opposite directions in the blue and red wings of the line profile, with the reversal occurring near the central wavelength of the line. These polarization structures can be discerned in earlier observations<sup>15,16</sup>, but have, hitherto, been left unexplained.

It is generally accepted that the broad Balmer lines in AGN are predominantly emitted by the accretion disk itself<sup>17</sup>. Polarization position angle rotations similar to those seen in PG 1700+518 are frequently observed in Seyfert 1 galaxies, in some of which the dominant polarization mechanism is believed to be scattering of emission line photons from the accretion disk in an equatorial scattering medium co-planar with, and closely surrounding, the disk. The characteristic  $\theta(\lambda)$  variation across the line profile naturally arises from the rotational motion of the disk, as long as the angular size of the disk as seen by the scattering medium is sufficiently large<sup>18</sup>. In the Seyfert galaxies, however, the scattered line profiles do not have the large

redshift seen in PG 1700+518. This net Doppler shift implies that the gas emitting the lines and the scattering electrons are moving apart with a velocity of  $\sim 4,000 \text{ km s}^{-1}$ .

In contrast to the Seyfert galaxies, therefore, the optical polarization spectrum of PG 1700+518 exhibits strong signatures of both rotational and radial motions. The most plausible interpretation of the redshifted scattered line profiles is that the scattering medium is part of a high velocity outflow. On the other hand, the rotation in  $\theta(\lambda)$  could be produced by rotational motions in the disk, or in the scattering region, or, as we argue below, both. Furthermore, the large



**Figure 1 | The polarization data for the BAL quasar PG1700+518.** The data were obtained using the William Herschel Telescope in conjunction with the ISIS spectrograph, at a spectral resolution of  $3.4 \text{ Å}$ , on the nights of 27 and 28 June 2003. The polarization data have been re-sampled into bins with an error of  $0.1\%$  in degree of polarization. **a**, The position angle of polarization,  $\theta$ ; **b**, the polarized flux spectrum in mJy; **c**, the degree of polarization as a percentage; and **d**, the total flux spectrum in mJy. Note the large change in polarization position angle across the broad H $\alpha$  emission line, which reverses direction at the line peak. In polarized flux, the broad H $\alpha$  emission line is redshifted with respect to the wavelength of the peak in total flux.

<sup>1</sup>Department of Physics, Rochester Institute of Technology, 84 Lomb Memorial Drive, Rochester, New York 14623, USA. <sup>2</sup>Centre for Astrophysics Research, Science & Technology Research Institute, University of Hertfordshire, Hatfield AL10 9AB, UK.



amplitude of the  $\theta(\lambda)$  swing implies that the scattering occurs close to the line-emitting region of the disk. The polarization spectrum, therefore, is consistent with scattering in a wind expanding away from the accretion disk. The observations cannot, however, be explained simply by adding a radial expansion component to the equatorial scattering region invoked to explain the broad H $\alpha$  polarization in Seyfert galaxies (effectively making it an equatorial wind).

We have used an extensive grid of three-dimensional scattering calculations, described in detail elsewhere<sup>19</sup>, to explore emission-line polarization signatures resulting from scattering in winds having different geometries, launched from a range of radii (relative to the Balmer-line emitting zone) within the disk. On quite general grounds, we infer that the scattering particles are electrons in ionized gas whose temperature is less than a few times  $10^5$  K. Scattering by dust grains can be ruled out, as the large  $\theta(\lambda)$  rotation requires that the scattering takes place within a radius of  $\sim 10^{16}$  m (0.3 pc) of the central AGN, well inside the radius ( $\sim 1$ – $2$  pc for this object) within which unshielded grains will be destroyed by the quasar's radiation field<sup>20</sup>. On the other hand, if the temperature of the scattering medium is  $\gg 10^5$  K, thermal Doppler broadening would smear out the polarization structure (Supplementary Fig. 1).

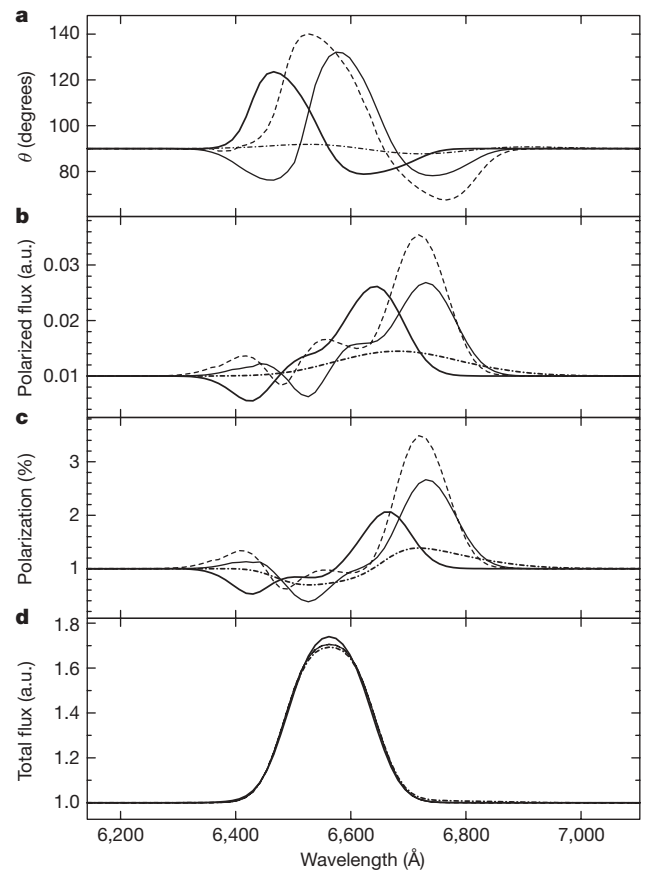
The model grid allows us to determine the essential elements such a scattering wind must possess in order to produce the principal features of the H $\alpha$  polarization, namely, a large amplitude rotation in  $\theta(\lambda)$ , anchored near the line centre, and the observed redshift in polarized flux. These features can be produced simultaneously only by models (Fig. 2, see also Supplementary Fig. 2) in which the wind rises vertically from the disk in a cylindrical configuration and includes both rotational ( $v_{\text{rot}}$ ) and vertical ( $v_z$ ) velocity components. Conical outflows naturally produce a redshifted line profile in polarized flux, but also a redshifted  $\theta(\lambda)$  rotation. For equatorial winds, on the other hand, the  $\theta(\lambda)$  rotation is both redshifted and of low amplitude. Furthermore, in order for the scattered line profile to have a Doppler width and redshift comparable to those observed, both the rotational and vertical velocity components of the cylindrical wind must have magnitudes comparable to the local keplerian velocity of the disk,  $v_K$  (Supplementary Fig. 3). The vertical extent ( $h$ ) of the scattering region above the disk is also strongly constrained by the observed polarization spectra. In order to produce a redshifted scattered line profile, while maintaining a large amplitude rotation in  $\theta(\lambda)$ , we require  $2r_{\text{in}} < h \leq 4r_{\text{in}}$ , where  $r_{\text{in}}$  is the inner radius of the scattering region in cylindrical coordinates (Supplementary Fig. 4). For similar reasons, the wind must be launched from approximately the same annular region within the disk as the Balmer-line emitting gas (Supplementary Fig. 5). Finally, values of  $p(\lambda)$  and the  $\theta(\lambda)$  rotation amplitude consistent with the observations are produced when the system is viewed at an intermediate inclination ( $i \approx 45^\circ$ ).

The radius of the Balmer-line emitting zone in PG 1700+518 has been determined by reverberation mapping to be  $\sim 7 \times 10^{15}$  m. Combining this radius with the Doppler velocity dispersion inferred from the width of the line profile leads (under the assumption that the emitting gas is gravitationally bound) to a mass for the central black hole of  $\sim 8 \times 10^8 M_\odot$  (ref. 21; here  $M_\odot$  is the solar mass). For this mass, the keplerian velocity at the launch radius of the wind is  $v_K \approx 4,000 \text{ km s}^{-1}$ . Our models imply that  $v_z \approx v_{\text{rot}} \approx v_K \approx 4,000 \text{ km s}^{-1}$ . This velocity is small compared to the outflow speeds characteristic of the BALs ( $1$ – $2 \times 10^4 \text{ km s}^{-1}$  in PG 1700+518; ref. 14), which implies that the scattering takes place in a region where the wind has not yet reached its terminal velocity, or where the acceleration mechanism is less efficient than in the BAL-producing zone.

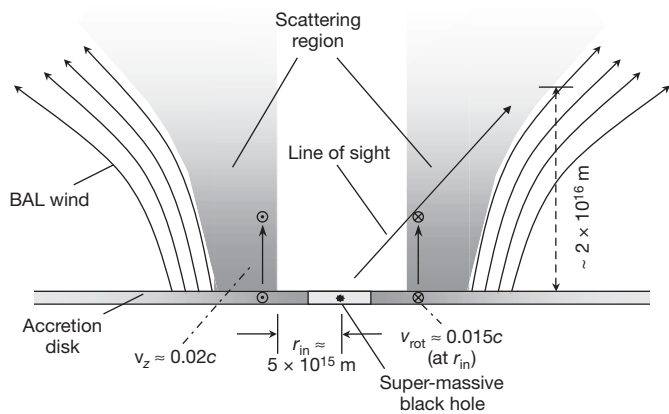
Theoretical studies have shown that disk winds driven either by radiation-pressure or magneto-centrifugal forces, or some combination thereof, are capable of producing velocities and mass-flow rates consistent with the observed properties of BALs<sup>22–24</sup>. Our interpretation of the spectropolarimetry data is consistent with a scenario in which the wind is launched from a relatively narrow annulus within the disk, which includes the Balmer-line emitting region, and is

initially accelerated vertically (Fig. 3). The velocity field at the base of the wind, with azimuthal and vertical components dominating the radial component, favours models in which the initial acceleration is due to thermal<sup>25</sup> or magnetic pressure<sup>26</sup> or radiation pressure of photons emitted locally by the disk<sup>10,27</sup>. On the other hand, our results are problematic for models invoking equatorial winds<sup>4,22</sup>, which require initial streamlines at large angles to the disk normal (for example,  $>30^\circ$  for a wind launched by magneto-centrifugal forces).

A viable radiatively driven disk wind must be shielded from the high energy continuum source of the AGN to prevent the gas becoming over-ionized, as this would eliminate line absorption, the major



**Figure 2 | Comparison of simulated polarization spectra for different scattering geometries.** The models were computed using our three-dimensional scattering code<sup>30</sup>, modified to include cylindrical and conical scattering winds. In each case, an emission line emitted by a quasar accretion disk is subsequently scattered in a wind. The wind is launched from the annular region of the accretion disk that produces the line emission. The models presented are: a cylindrical wind with both rotational and vertical outflow velocity components (solid heavy line); a disk-like, rotating, equatorial outflow (dot-dash heavy line); and a hollow-funnel-like conical wind with radial outflow only (solid lighter line) and with radial outflow combined with rotation (dashed lighter line). The inclination of the system polar axis to the line of sight is  $48^\circ$ , the outflow velocity is  $0.02c$  and where applicable, the winds are rotating at the local keplerian velocity. The position angle of polarization ( $\theta$ ; **a**) is measured relative to the projection of the system axis on the sky plane. The median  $\theta$  for the equatorial outflow model is  $0^\circ$ , but has been shifted by  $90^\circ$  to ease comparison with the other models. The total flux spectrum (**d**) and the polarized flux spectrum (**b**) are in arbitrary units (a.u.); the degree of polarization (%) is shown in **c**. The equatorial outflow produces a red-shifted broad line in polarized flux but only a small variation in  $\theta$  across the line. Both the cylindrical and conical wind models produce large-amplitude rotations of  $\theta$ , but only for the cylindrical wind is the rotation centred at the peak of the total flux line profile, as required. A colour version of this figure is included as Supplementary Information.



**Figure 3 | A cross-section schematic illustration of the wind geometry inferred from the polarization and spectroscopic properties of PG1700+518.** The scattering wind launches vertically from the accretion disk over a range of radii approximately coincident with the region emitting the bulk of the Balmer line emission. The emission disk and wind are rotating in the sense indicated by the symbols (dotted circle, motion out of the plane of the diagram; crossed circle, motion into the plane of the diagram). Radiation emitted by the accretion disk is scattered in the cylindrical inner section of the wind (shaded), which rises to  $\sim 4r_{\text{in}}$ , where  $r_{\text{in}}$  is the inner radius of the broad emission-line region. The rotational and vertical components of the wind's velocity are both comparable with the local keplerian velocity. The broad absorption lines (BALs) are formed in the outer section of the wind, which is shielded from the high-energy radiation of the central AGN continuum by the scattering wind. The acceleration of this outer wind is dominated by line absorption of ultraviolet radiation from the central source, causing the flow to diverge along radial streamlines.

source of the radiation pressure driving force<sup>22</sup>. Theoretical work on this problem suggests that disk winds are self-shielded by an inner zone, which is too highly ionized for efficient line driving but which is effectively opaque to X-rays<sup>24,27</sup>. The scattering region can be plausibly identified with such an X-ray-shielding inner wind zone. Our polarimetry data do not directly constrain the region in which the BALs arise. However, in such two-zone wind models, these features are formed in the shielded outer wind<sup>27</sup>. Here radiative acceleration due to ultraviolet line opacity is efficient, and the streamlines, although nearly vertical close to the disk, become effectively radial at higher elevations (Fig. 3).

It has been argued that disk winds explain many observed properties of AGN, including BALs, the X-ray 'warm absorber' and ultraviolet absorption lines<sup>28</sup>. The spectropolarimetric properties of PG 1700+518 provide compelling evidence not only that such disk winds exist but also that they are a key element in any unified picture of the inner structure of AGN. More specifically, our results demonstrate that the polarization structure across the broad emission lines constrains the geometry, launch site and kinematics of the scattering medium, and therefore places vital and unique observational constraints on physical models of disk winds. Furthermore, these insights into the nature of AGN disk winds may also be applicable to a wide variety of other astrophysical sources in which keplerian disks are thought to generate outflows, such as young stellar objects, or B[e] stars. One of the latter, in particular, exhibits emission line polarization structures reminiscent of those observed in PG 1700+518 (ref. 29).

Received 21 March; accepted 20 September 2007.

1. Rees, M. J. Black hole models for active galactic nuclei. *Annu. Rev. Astron. Astrophys.* **22**, 471–506 (1984).
2. Ferrarese, L. & Ford, H. Supermassive black holes in galactic nuclei: Past, present and future research. *Space Sci. Rev.* **116**, 523–624 (2005).

3. Crenshaw, D. M., Kraemer, S. B. & George, I. M. Mass loss from the nuclei of active galaxies. *Annu. Rev. Astron. Astrophys.* **41**, 117–167 (2003).
4. Blandford, R. D. & Payne, D. G. Hydromagnetic flows from accretion discs and the production of radio jets. *Mon. Not. R. Astron. Soc.* **199**, 883–903 (1982).
5. Silk, J. & Rees, M. J. Quasars and galaxy formation. *Astron. Astrophys.* **331**, L1–L4 (1998).
6. Scannapieco, E., Silk, J. & Bouwens, R. AGN feedback causes downsizing. *Astrophys. J.* **635**, L13–L16 (2005).
7. Di Matteo, T., Springel, V. & Hernquist, L. Energy input from quasars regulates the growth and activity of black holes and their host galaxies. *Nature* **433**, 604–607 (2005).
8. Bower, R. G. *et al.* Breaking the hierarchy of galaxy formation. *Mon. Not. R. Astron. Soc.* **370**, 645–655 (2006).
9. Menci, N., Fontana, A., Giallongo, E., Grazian, A. & Salimbeni, S. The abundance of distant and extremely red galaxies: The role of AGN feedback in hierarchical models. *Astrophys. J.* **647**, 753–762 (2006).
10. Shlosman, I., Vitello, P. A. & Shaviv, G. Active galactic nuclei — Internal dynamics and formation of emission clouds. *Astrophys. J.* **294**, 96–105 (1985).
11. Turnshek, D. A. Properties of the broad absorption-line QSOs. *Astrophys. J.* **280**, 51–65 (1984).
12. Reichard, T. A. *et al.* Continuum and emission-line properties of broad absorption line quasars. *Astron. J.* **126**, 2594–2607 (2003).
13. Weymann, R. J., Morris, S. L., Foltz, C. B. & Hewett, P. C. Comparisons of the emission-line and continuum properties of broad absorption line and normal quasi-stellar objects. *Astrophys. J.* **373**, 23–53 (1991).
14. Pettini, M. & Boksenberg, A. PG 1700+518 — A low-redshift, broad absorption line QSO. *Astrophys. J.* **294**, L73–L78 (1985).
15. Schmidt, G. D., Hines, D. C., The polarization of broad absorption line QSOs. *Astrophys. J.* **512**, 125–135 (1999).
16. Ogle, P. M. *et al.* Polarization of broad absorption line QSOs. I. A spectropolarimetric atlas. *Astrophys. J. Suppl. Ser.* **125**, 1–34 (1999).
17. Rokaki, E. & Boisson, C. Consistency of accretion discs with Seyfert 1 UV fluxes and H $\beta$  emission-line profiles. *Mon. Not. R. Astron. Soc.* **307**, 41–54 (1999).
18. Smith, J. E., Robinson, A., Young, S., Axon, D. J. & Corbett, E. A. Equatorial scattering and the structure of the broad-line region in Seyfert nuclei: Evidence for a rotating disc. *Mon. Not. R. Astron. Soc.* **359**, 846–864 (2005).
19. Young, S., Robinson, A., Axon, D. J. & Hough, J. H. Polarization signatures of disc-winds. *Astrophys. J.* (submitted).
20. Barvainis, R. Hot dust and the near-infrared bump in the continuum spectra of quasars and active galactic nuclei. *Astrophys. J.* **320**, 537–544 (1987).
21. Peterson, B. M. *et al.* Central masses and broad-line region sizes of active galactic nuclei. II. A homogeneous analysis of a large reverberation-mapping database. *Astrophys. J.* **613**, 682–699 (2004).
22. Murray, N., Chiang, J., Grossman, S. A. & Voit, G. M. Accretion disk winds from active galactic nuclei. *Astrophys. J.* **451**, 498–509 (1995).
23. de Kool, M. & Begelman, M. C. Radiation pressure-driven magnetic disk winds in broad absorption line quasi-stellar objects. *Astrophys. J.* **455**, 448–455 (1995).
24. Proga, D., Stone, J. M. & Kallman, T. R. Dynamics of line-driven disk winds in active galactic nuclei. *Astrophys. J.* **543**, 686–696 (2000).
25. Begelman, M. C., McKee, C. F. & Shields, G. A. Compton heated winds and coronae above accretion disks. I Dynamics. *Astrophys. J.* **271**, 70–88 (1983).
26. Stone, J. M. & Norman, M. L. Numerical simulations of magnetic accretion disks. *Astrophys. J.* **433**, 746–756 (1994).
27. Proga, D. & Kallman, T. R. Dynamics of line-driven disk winds in active galactic nuclei. II. Effects of disk radiation. *Astrophys. J.* **616**, 688–695 (2004).
28. Elvis, M. A Structure for quasars. *Astrophys. J.* **545**, 63–76 (2000).
29. Oudmaijer, R. D., Proga, D., Drew, J. E. & de Winter, D. The evolved B[e] star HD 87643: Observations and radiation driven disk-wind model for B[e] stars. *Mon. Not. R. Astron. Soc.* **300**, 170–182 (1998).
30. Young, S. A generic scattering model for AGN. *Mon. Not. R. Astron. Soc.* **312**, 567–578 (2000).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work is based on observations made with the William Herschel Telescope operated on the island of La Palma by the Isaac Newton Group in the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofísica de Canarias. This research has made use of NASA's Astrophysics Data System. We acknowledge financial support from the Science and Technology Facilities Council, UK.

**Author Contributions** All authors contributed extensively to the work presented in this paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to S.Y. (sxysps@rit.edu).

# First-passage times in complex scale-invariant media

S. Condamin<sup>1</sup>, O. Bénichou<sup>1</sup>, V. Tejedor<sup>1</sup>, R. Voituriez<sup>1</sup> & J. Klafter<sup>2</sup>

How long does it take a random walker to reach a given target point? This quantity, known as a first-passage time (FPT), has led to a growing number of theoretical investigations over the past decade<sup>1</sup>. The importance of FPTs originates from the crucial role played by first encounter properties in various real situations, including transport in disordered media<sup>2,3</sup>, neuron firing dynamics<sup>4</sup>, spreading of diseases<sup>5</sup> or target search processes<sup>6–9</sup>. Most methods of determining FPT properties in confining domains have been limited to effectively one-dimensional geometries, or to higher spatial dimensions only in homogeneous media<sup>1</sup>. Here we develop a general theory that allows accurate evaluation of the mean FPT in complex media. Our analytical approach provides a universal scaling dependence of the mean FPT on both the volume of the confining domain and the source–target distance. The analysis is applicable to a broad range of stochastic processes characterized by length-scale-invariant properties. Our theoretical predictions are confirmed by numerical simulations for several representative models of disordered media<sup>10</sup>, fractals<sup>3</sup>, anomalous diffusion<sup>11</sup> and scale-free networks<sup>12</sup>.

Transport properties are often characterized by the exit time from a sphere  $t_{\text{exit}}$ , which is the first time a random walker reaches any point at a distance  $r$  from its starting point. This quantity is well known for brownian motion in euclidean spaces, and has also been evaluated for finitely ramified deterministic fractals<sup>13,14</sup>. In these cases, the length-scale-invariant properties of the walker's trajectories have a key role and lead to the scaling form  $t_{\text{exit}} \propto r^{d_w}$ , which defines the walk dimension<sup>3</sup>  $d_w$ . Interestingly, it has been shown that a large class of complex scale-free networks are also invariant under the length scale renormalization scheme defined in ref. 15, even if they are of 'small world' type—that is, if their diameter scales like the logarithm of the volume. This remarkable property led the authors of ref. 12 in particular to characterize the mean exit time in this class of small-world networks by a set of scaling exponents.

However, in many situations, the determining quantity is not  $t_{\text{exit}}$ , but rather the FPT of a random walk starting from a source point  $S$  to a given target point  $T$ . Indeed the FPT is a key quantity to quantify the kinetics of transport-limited reactions<sup>14,16</sup>, which encompass not only chemical or biochemical reactions<sup>17,18</sup>, but also (at larger scales) interactions involving more complex organisms, such as a virus infecting a cell<sup>19</sup> or animals searching for food<sup>6</sup>. The relevance of the FPT has also been recently highlighted in ref. 12 in the context of scale-free networks, such as social networks<sup>20</sup>, protein interaction networks<sup>21</sup> or metabolic networks<sup>22</sup>. The FPT and the exit time in fact possess very different properties. Indeed, the exit time is not sensitive to the confinement, as only a sphere of radius  $r$  is explored by the random walker. On the contrary, an estimation of the time needed to go from one point to another, namely the FPT, crucially depends on the confining environment—the mean FPT (MFPT) being actually infinite in unbounded domains.

Consider a random walker moving in a bounded domain of size  $N$ . Let  $W(\mathbf{r}, t | \mathbf{r}')$  be the propagator (that is, the probability density of being at site  $\mathbf{r}$  at time  $t$ , starting from site  $\mathbf{r}'$  at time 0), and  $P(\mathbf{r}, t | \mathbf{r}')$

the probability density that the FPT to reach  $\mathbf{r}$ , starting from  $\mathbf{r}'$ , is  $t$ . These two probability densities are known to be related through<sup>23</sup>

$$W(\mathbf{r}_T, t | \mathbf{r}_S) = \int_0^t P(\mathbf{r}_T, t' | \mathbf{r}_S) W(\mathbf{r}_T, t - t' | \mathbf{r}_T) dt' \quad (1)$$

where  $\mathbf{r}_S$  and  $\mathbf{r}_T$  denote, respectively, the source and target position. After integration over  $t$ , this equation gives an exact expression for the MFPT, provided it is finite:

$$\langle T \rangle = \frac{H(\mathbf{r}_T | \mathbf{r}_T) - H(\mathbf{r}_T | \mathbf{r}_S)}{W_{\text{stat}}(\mathbf{r}_T)} \quad (2)$$

where

$$H(\mathbf{r} | \mathbf{r}') = \int_0^\infty (W(\mathbf{r}, t | \mathbf{r}') - W_{\text{stat}}(\mathbf{r})) dt \quad (3)$$

and  $W_{\text{stat}}$  is the stationary probability distribution (see Supplementary Information for details). Equation (2) is an extension of an analogous form given in ref. 24, for which no quantitative determination of the MFPT could be proposed. The main problem at this stage is to determine the unknown function  $H$ , which is indeed a complicated task, as it depends both on the walk's characteristics and on the shape of the domain. A crucial step that allows us to go further in the general case is that  $H$  turns out to be the pseudo-Green function of the domain<sup>25</sup>, which in turn is well suited to a quantitative analysis. Indeed, we propose approximating  $H$  by its infinite-space limit, which is precisely the usual Green function  $G_0$ :

$$H(\mathbf{r} | \mathbf{r}') \approx G_0(\mathbf{r} | \mathbf{r}') = \int_0^\infty W_0(\mathbf{r}, t | \mathbf{r}') dt \quad (4)$$

where  $W_0$  is the infinite space propagator (Supplementary Information). Note that a similar approximation has proven to be satisfactory in the standard example of regular diffusion<sup>26</sup>. We stress that when inserted in equation (2), this form does not lead to a severe infinite space approximation of the MFPT, because all the dependence on the domain geometry is now contained in the factor  $1/W_{\text{stat}}$ . This approximation is the key step of our derivation and, as we proceed to show, captures extremely well the confining effects on MFPTs in complex media.

We first consider the case of a uniform stationary distribution  $W_{\text{stat}} = 1/N$ , which is realized as soon as the links of the network are not directed and the number of connected neighbours of a node, the degree, is constant. This assumption amounts to symmetrical transition rates and actually underlies many models of transport in complex media, with the notable exception of scale-free networks, which will be tackled later on in this Letter. Following ref. 3, we assume for  $W_0$  the standard scaling:

$$W_0(\mathbf{r}, t | \mathbf{r}') \propto t^{-d_t/d_w} \Pi\left(\frac{|\mathbf{r} - \mathbf{r}'|}{t^{1/d_w}}\right) \quad (5)$$

<sup>1</sup>Université Pierre et Marie Curie-Paris 6, Laboratoire de Physique Théorique de la Matière Condensée, UMR CNRS 7600, case 121, 4 Place Jussieu, 75005 Paris, France. <sup>2</sup>School of Chemistry, Tel Aviv University, Tel Aviv 69978, Israel.



where the fractal dimension  $d_f$  characterizes the number of sites  $N_r \propto r^{d_f}$  within a sphere of radius  $r$ ,  $\Pi$  is the infinite space scaling function, and  $d_w$  has been defined previously. This form ensures the normalization of  $W_0$  by integration over the whole fractal set. A derivation given in Supplementary Information then yields our central result:

$$\langle T \rangle \sim \begin{cases} N(A - Br^{d_w - d_f}) & \text{for } d_w < d_f \\ N(A + B \ln r) & \text{for } d_w = d_f \\ N(A + Br^{d_w - d_f}) & \text{for } d_w > d_f \end{cases} \quad (6)$$

for  $r = |\mathbf{r}_T - \mathbf{r}_S|$  different from 0; here ' $\sim$ ' indicates large  $N$  asymptotic equivalence. Strikingly, the constants  $A$  and  $B$  do not depend on the confining domain. In addition, whereas  $A$  is related to the small-scale properties of the walk, we emphasize that  $B$  can be written solely in terms of  $\Pi$  (a precise definition of  $A$  and  $B$  is given in Supplementary Information). These expressions therefore unveil a universal scaling dependence of the MFPT on the geometrical parameters  $N$  and  $r$ .

Several comments are in order. First, we point out that equation (6) gives the large  $N$  asymptotics of the MFPT, which is a function of  $N$  and  $r$  as independent variables. In particular, the volume dependence is linear in  $N$  for  $r$  fixed in any case, which can not be inferred from the standard scaling  $\langle T \rangle \propto L^{d_w}$ ,  $L$  being the characteristic length of the domain of order  $N^{1/d_f}$ . Second, a global rescaling of the problem  $r \rightarrow \lambda r$ ,  $L \rightarrow \lambda L$ , when applied to equation (6), gives the standard form  $\langle T \rangle \propto \lambda^{d_w}$  for  $d_w > d_f$  and  $\langle T \rangle \propto \lambda^{d_f}$  for  $d_w < d_f$  in accord with refs 27 and 28. Third, equation (6) shows two regimes, which rely on infinite space properties of the walk: in the case of compact exploration<sup>3</sup> ( $d_w \geq d_f$ ) where each site is eventually visited, the MFPT behaves like  $\langle T \rangle \propto Nr^{d_w - d_f}$  ( $\langle T \rangle \propto N \ln r$  for  $d_w = d_f$ ) at large distances, so that the dependence on the starting point always matters; in the opposite case of non-compact exploration,  $\langle T \rangle$  tends to a finite value for large  $r$ , and the dependence on the starting point is lost.

We now confirm these analytical results by Monte Carlo simulations and exact enumeration methods applied to various models that exemplify the three previous cases. (1) The random barrier model in two dimensions<sup>3</sup> is a widespread model of transport in disordered systems in which MFPT properties remain widely unexplored. It is defined by a lattice random walk with nearest neighbours symmetrical transition rates  $\Gamma$  distributed according to some distribution  $\rho(\Gamma)$ . Even for a power law distribution  $\rho(\Gamma)$  the scaling function  $\Pi(\xi)$  can be shown to be gaussian<sup>10</sup> ( $d_f = d_w = 2$ ), which allows us to explicitly compute the constant  $B$  and obtain  $\langle T \rangle \sim N(A + (1/2\pi D_{\text{eff}}) \ln r)$ . Here  $D_{\text{eff}}$  is a diffusion constant depending on  $\rho(\Gamma)$  that can be determined by an effective medium approximation<sup>10</sup> (Supplementary

Information). (2) The Sierpinski gasket of finite order is a representative example of deterministic fractals, described in Fig. 1. In this case<sup>3</sup>  $d_f = \ln 3 / \ln 2 < \ln 5 / \ln 2 = d_w$ , so that our theory predicts the scaling  $\langle T \rangle \propto Nr^{(\ln 5 - \ln 3) / \ln 2}$ . (3) The Lévy flight model of anomalous diffusion<sup>11,23</sup> is based on a fat-tailed distribution of jump lengths  $p(l) \propto l^{-d-\beta}$  ( $0 < \beta \leq 2$ ). The walk dimension is now  $d_w = \beta$ , whereas the fractal dimension is the dimension of the euclidian space  $d$ . In dimensions  $d \geq 2$ , or for  $d = 1$  when  $\beta < 1$ , one has  $d_f > d_w$  and our theory gives  $\langle T \rangle \sim N(A - Br^{\beta-d})$ .

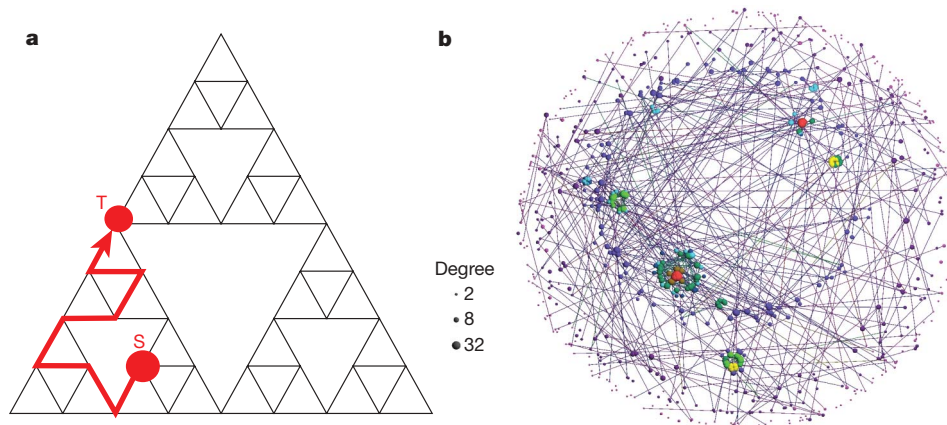
Figure 2a–c reveals excellent quantitative agreement between the analytical predictions and the numerical simulations. Both the volume dependence and the source–target distance dependence are unambiguously captured by our theoretical expressions, equation (6), as shown by the data collapse of the numerical simulations. We emphasize that the very different nature of these examples demonstrates that the range of applicability of our approach, which mainly relies on the length-scale-invariant property of the infinite space propagator (equation (5)), is wide.

These analytical results can be extended to scale-free networks. The latter are characterized by a power-law degree distribution. A wide class of scale-free networks has been proven recently to be invariant under a length-scale renormalization scheme defined in ref. 15: social networks<sup>20</sup>, the world wide web<sup>29</sup>, metabolic networks<sup>22</sup>, and yeast protein interaction networks (PIN)<sup>21</sup>. Although the standard fractal dimension  $d_f$  of these networks is infinite as their diameter scales as  $\ln N$ , one can show that they are scale-invariant in the following sense: they can be covered with  $N_B$  non-overlapping boxes of size  $l_B$  with  $N_B/N \propto l_B^{-d_B}$ . This renormalization property defines an alternative scaling exponent called the box dimension  $d_B$ , which is actually equal to the fractal dimension defined earlier as long as the networks are not of small-world type. A model of scale-free networks possessing such length-scale-invariant properties has been defined recursively in refs 12 and 30: the network grows by adding  $m$  new offspring nodes to each existing network node, resulting in well defined modules. In addition, modules are connected to each other through  $x$  random links (Supplementary Information). In this case  $d_B = \ln(2m + x) / \ln 3$  and  $d_w = \ln(6m/x + 3) / \ln 3$ .

For this class of networks,  $W_{\text{stat}}(\mathbf{r})$  is not uniform any more but proportional to the degree  $k(\mathbf{r})$  of the node  $\mathbf{r}$ . One can use the length-scale-invariant property to infer the following scaling of the infinite space propagator:

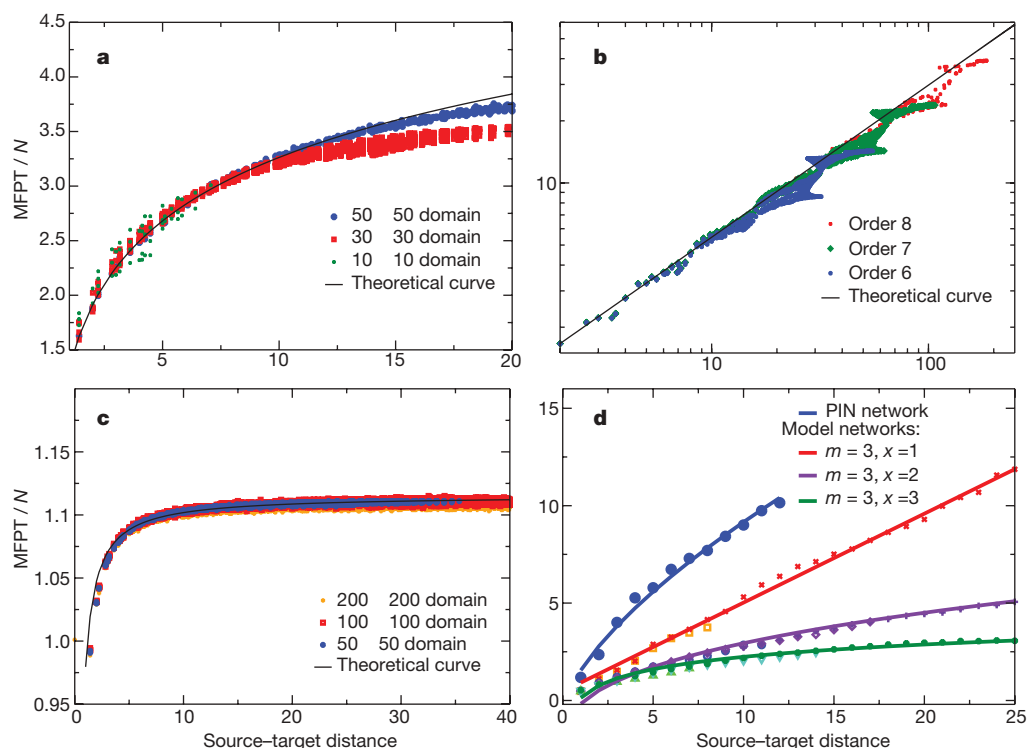
$$\frac{W_0(\mathbf{r}, t | \mathbf{r}')}{k(\mathbf{r})} \propto t^{-d_B/d_w} \Pi\left(\frac{|\mathbf{r} - \mathbf{r}'|}{t^{1/d_w}}\right) \quad (7)$$

This form, compatible with the symmetry relations proposed in ref. 24, allows us to perform a similar derivation, which leads for the MFPT to the same result (equation (6)), but where  $d_f$  is to be replaced



**Figure 1 | Length-scale-invariant networks.** **a**, The Sierpinski gasket (here of order three) is a representative example of a deterministic fractal. A sample random path from  $S$  to  $T$  is shown. **b**, The yeast PIN, obtained from

the filtered yeast interactome developed in ref. 21. Picture generated by LaNet-vi software (<http://xavier.informatics.indiana.edu/lanet-vi/>).



**Figure 2 | Mean FPT in complex media.** **a**, Random barrier model with a transition rate distribution  $\rho(J) = (\alpha/J)(J/\Gamma_0)^\alpha$ , with  $\Gamma_0 = 1$  and  $\alpha = 0.5$ . The confining domain is an  $L \times L$  square with the target point in the middle. Shown are numerical simulations of the MFPT rescaled by the volume  $N$ , averaged over the disorder, for three different domain sizes. The theoretical curve (black line) is given by  $\langle T \rangle / N \sim (A + (1/2\pi D_{\text{eff}}) \ln r)$ , where the only fitting parameter is  $A$ ;  $D_{\text{eff}}$  is evaluated in Supplementary Information. **b**, Numerical simulations of random walks on a Sierpinski gasket (log/log plot) for three different system sizes (orders 6, 7 and 8). For each set of points, the size of the Sierpinski gasket and the target point are fixed (the target point corresponds to the point T on the Sierpinski gasket of order 3 in Fig. 1a, and the starting point takes various positions on the Sierpinski gasket). The black line corresponds to the theoretical scaling  $r^{d_B - d_L}$ . **c**, Simulations of Lévy flights on a two-dimensional square lattice ( $\beta = 1$ ). The confining domains are  $50 \times 50$ ,  $100 \times 100$  and  $200 \times 200$  squares, with

the target in the middle. The MFPT is presented as a function of the source–target distance for different source points. Simulation points are fitted with  $\langle T \rangle / N \sim (A - Br^{\beta-2})$ . **d**, Simulations of random walks on fractal complex networks of small-world type. The MFPT on the PIN network (blue circles) is fitted by  $\langle T \rangle / N \sim (A + Br^{d_w - d_B})$  with  $d_w \approx 2.86$  and  $d_B \approx 2.2$ , as found in ref. 12. We also consider three examples of the model of networks defined in ref. 12: ( $m = 3$ ,  $x = 1$ ,  $d_B - d_w = 1$ ), red symbols and fitting curve; ( $m = 3$ ,  $x = 2$ ,  $d_B - d_w = \ln(3/2)/\ln 3$ ), violet symbols and fitting curve; and ( $m = 3$ ,  $x = 3$ ,  $d_B - d_w = 0$ ), green symbols and fitting curve. For each example, the MFPT (rescaled by the network volume  $N = (1+x)^k$  for three system sizes  $k = 3, 4, 5$ ) averaged over the disorder is presented as a function of the source–target distance for different source points, and fitted by the theoretical expression  $\langle T \rangle / N \sim (A + Br^{\ln(3/x)/\ln 3})$ . We find quite surprisingly a scaling independent of  $m$ .

by  $d_B$ . We applied this formula to an example of a scale-free biological network, the yeast PIN (Fig. 1b), obtained from the filtered yeast interactome developed in ref. 21, and to a model<sup>12,30</sup> of a scale-free fractal network. Figure 2d shows that this analytical result is in good agreement with numerical simulations on the PIN network. The data collapse over various system sizes for the model of scale-free fractal networks provides a further validation of our approach, and indicates that our theory has a wide range of applications.

Received 28 June; accepted 23 August 2007.

- Redner, S. *A Guide to First-Passage Processes* (Cambridge Univ. Press, Cambridge, UK, 2001); Errata (<http://physics.bu.edu/~redner/>).
- Havlin, S. & ben-Avraham, D. Diffusion in disordered media. *Adv. Phys.* **36**, 695–798 (1987).
- ben-Avraham, D. & Havlin, S. *Diffusion and Reactions in Fractals and Disordered Systems* (Cambridge Univ. Press, Cambridge, UK, 2000).
- Tuckwell, H. C. *Introduction to Theoretical Neurobiology* (Cambridge Univ. Press, Cambridge, UK, 1988).
- Lloyd, A. L. & May, R. M. Epidemiology — how viruses spread among computers and people. *Science* **292**, 1316–1317 (2001).
- Bénichou, O., Coppey, M., Moreau, M., Suet, P. H. & Voituriez, R. Optimal search strategies for hidden targets. *Phys. Rev. Lett.* **94**, 198101 (2005).
- Bénichou, O., Loverdo, C., Moreau, M. & Voituriez, R. Two-dimensional intermittent search processes: An alternative to Lévy flight strategies. *Phys. Rev. E* **74**, 020102 (2006).
- Shlesinger, M. F. Mathematical physics: Search research. *Nature* **443**, 281–282 (2006).

- Eliazar, I., Koren, T. & Klafter, J. Searching circular DNA strands. *J. Phys. Condens. Matter* **19**, 065140 (2007).
- Bouchaud, J.-P. & Georges, A. Anomalous diffusion in disordered media: statistical mechanisms, models and applications. *Phys. Rep.* **195**, 127–293 (1990).
- Metzler, R. & Klafter, J. The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Phys. Rep.* **339**, 1–77 (2000).
- Gallos, L. K., Song, C., Havlin, S. & Makse, H. A. Scaling theory of transport in complex biological networks. *Proc. Natl Acad. Sci. USA* **104**, 7746–7751 (2007).
- Van den Broeck, C. Renormalization of first-passage times for random walks on deterministic fractals. *Phys. Rev. A* **40**, 7334–7345 (1989).
- Yuste, S. B. First-passage time, survival probability and propagator on deterministic fractals. *J. Phys. A* **28**, 7027–7038 (1995).
- Song, C., Havlin, S. & Makse, H. A. Self-similarity of complex networks. *Nature* **443**, 392–395 (2005).
- Rice, S. A. *Diffusion-Limited Reactions* (Elsevier, Amsterdam, 1985).
- Berg, O. G., Winter, R. B. & von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **20**, 6929–6948 (1981).
- Coppey, M., Bénichou, O., Voituriez, R. & Moreau, M. Kinetics of target site localization of a protein on DNA: A stochastic approach. *Biophys. J.* **87**, 1640–1649 (2004).
- Holcman, D. Modeling DNA and virus trafficking in the cell cytoplasm. *J. Stat. Phys.* **127**, 471–494 (2007).
- Barabási, A. L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
- Han, J.-D. J. et al. Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature* **430**, 88–93 (2004).

22. Almaas, E., Kovacs, B., Vicsek, T., Oltvai, Z. N. & Barabasi, A. L. Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature* **427**, 839–843 (2004).
23. Hughes, B. D. *Random Walks and Random Environments* (Oxford Univ. Press, New York, 1995).
24. Noh, J. D. & Rieger, H. Random walks on complex networks. *Phys. Rev. Lett.* **92**, 118701 (2004).
25. Barton, G. *Elements of Green Functions and Propagation: Potentials, Diffusion and Waves* (Oxford Univ. Press, New York, 1989).
26. Condamin, S., Bénichou, O. & Moreau, M. First-passage times for random walks in bounded domains. *Phys. Rev. Lett.* **95**, 260601 (2005).
27. Montroll, E. W. Random walks on lattices. iii. Calculation of first-passage times with application to exciton trapping on photosynthetic units. *J. Math. Phys.* **10**, 753–765 (1969).
28. Boltt, E. M. & ben-Avraham, D. What is special about diffusion on scale-free nets? *N. J. Phys.* **7**, 26–47 (2005).
29. Albert, R., Jeong, H. & Barabasi, A. L. Internet: Diameter of the world-wide web. *Nature* **401**, 130–131 (1999).
30. Song, C., Havlin, S. & Makse, H. A. Origins of fractality in the growth of complex networks. *Nature Phys.* **2**, 275–281 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to J. M. Victor for discussions.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to O.B. ([benichou@lptmc.jussieu.fr](mailto:benichou@lptmc.jussieu.fr)).



# Abrupt onset of a second energy gap at the superconducting transition of underdoped Bi2212

W. S. Lee<sup>1</sup>, I. M. Vishik<sup>1</sup>, K. Tanaka<sup>1,2</sup>, D. H. Lu<sup>1</sup>, T. Sasagawa<sup>1</sup>, N. Nagaosa<sup>3</sup>, T. P. Devereaux<sup>4</sup>, Z. Hussain<sup>2</sup> & Z.-X. Shen<sup>1</sup>

The superconducting gap—an energy scale tied to the superconducting phenomena—opens on the Fermi surface at the superconducting transition temperature ( $T_c$ ) in conventional BCS superconductors. In underdoped high- $T_c$  superconducting copper oxides, a pseudogap (whose relation to the superconducting gap remains a mystery) develops well above  $T_c$  (refs 1, 2). Whether the pseudogap is a distinct phenomenon or the incoherent continuation of the superconducting gap above  $T_c$  is one of the central questions in high- $T_c$  research<sup>3–8</sup>. Although some experimental evidence suggests that the two gaps are distinct<sup>9–18</sup>, this issue is still under intense debate. A crucial piece of evidence to firmly establish this two-gap picture is still missing: a direct and unambiguous observation of a single-particle gap tied to the superconducting transition as function of temperature. Here we report the discovery of such an energy gap in underdoped  $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$  in the momentum space region overlooked in previous measurements. Near the diagonal of Cu–O bond direction (nodal direction), we found a gap that opens at  $T_c$  and has a canonical (BCS-like) temperature dependence accompanied by the appearance of the so-called Bogoliubov quasi-particles, a classical signature of superconductivity. This is in sharp contrast to the pseudogap near the Cu–O bond direction (antinodal region) measured in earlier experiments<sup>19–21</sup>.

In spectra taken by angle-resolved photoemission spectroscopy (ARPES), the temperature dependence of the gap seems to be very different between the nodal region and the antinodal region, as demonstrated by the Fermi-function divided spectra of the UD92K sample shown in Fig. 1a. Above  $T_c$ , there is a gapless Fermi arc near the nodal region (C1 to C3 at 120 K); away from this Fermi arc region, the well-known pseudogap gradually takes over and reaches its maximum at the antinodal region (C5 to C8 at 120 K)<sup>1,2,19–22</sup>. Below  $T_c$ , the magnitude of the pseudogap at the antinode (C8 at 82 K and 10 K) does not show clear temperature dependence across  $T_c$ , although a sharper peak in the spectrum develops in the superconducting state, as shown in Fig. 1c. The lack of temperature dependence of the antinodal gap size is well known and has been taken as evidence of pairing above  $T_c$  (ref. 3). In contrast, along the Fermi surface near the nodal region (C1 to C3), an energy gap opens up just below  $T_c$  and becomes larger as the system cools to a temperature well below  $T_c$ . We note that at 82 K there is appreciable thermal population above the Fermi energy ( $E_F$ ) such that the upper branch of the Bogoliubov dispersion can be clearly seen in the raw spectra for C1 to C4, as shown in Fig. 1b. This observation demonstrates that the near-nodal gap is related to superconductivity, because Bogoliubov quasi-particles exist only in the coherent superconducting state.

The temperature dependence of the gap evolution near the nodal region of the UD92K sample is analysed and demonstrated in Fig. 2, suggesting that  $T_c$  is correlated with the opening of a single particle gap. Starting with the common procedure of using the symmetrized energy distribution curves (EDCs) at the Fermi crossing point, the data in Fig. 2c suggest a collapse of the superconducting gap very close to  $T_c$ . As shown in Fig. 2d, the extracted gap size at locations A, B and C is obtained by fitting the symmetrized EDCs to a phenomenological model<sup>23</sup> containing a minimal set of parameters: the gap size and the lifetime broadening of the quasi-particles (see Supplementary Information). As can be seen, the gap at these three locations gradually closes as the temperature approaches  $T_c$  and vanishes at a temperature close to  $T_c$ , following the functional form of  $\Delta(T)$  in weak-coupling BCS theory surprisingly well. The second indication of gap closing comes from the Bogoliubov quasi-particle dispersion as shown in Fig. 2a, b. The shift of the Bogoliubov peak suggests that the superconducting gap size decreases rapidly when the temperature approaches  $T_c$ . Further, the disappearance of the Bogoliubov band within a narrow temperature range (87–97 K) having small thermal broadening difference confirms the collapse of the gap above  $T_c$  (see Supplementary Information). Third, we remark that this sudden onset of the gap is closely related to the abrupt decrease in the quasi-particle lifetime at  $T_c$ , as shown in the fitted  $\Gamma$  from our data (Supplementary Fig. 2b). This observation is consistent with microwave spectroscopy<sup>24</sup>, thermal Hall conductivity<sup>25</sup> and ARPES data of nodal quasi-particles<sup>26,27</sup>. Thus, the temperature dependence of the electronic states in the nodal region is consistent with that of a BCS superconductor.

We note that such behaviour has not been observed in previous ARPES measurements<sup>23</sup> because this nodal ‘BCS-gap’ region has been overlooked as a result of insufficient momentum-space sampling. The gap at previously reported momentum positions has already opened above  $T_c$  because of the proximity to the pseudogap region, whose behaviour was extrapolated to the entire Fermi surface<sup>20,23</sup>, leading to the conclusion that the pseudogap and the superconducting gap were one and the same. Our finding indicates that this extrapolation and its conclusion should be revised.

To study the temperature evolution of the gap function, spectra at three selected temperatures were recorded to accommodate a large number of cuts along the Fermi surface. As shown in Fig. 3a, the gap along the Fermi surface can be divided roughly into two groups. One group is the region near the node (C1 to C2) where the gap is temperature dependent with an onset temperature very close to  $T_c$ . The other group is associated with the antinodal region (C6 to C7), which does not show any significant temperature dependence across  $T_c$ . As we move towards the antinode (including curves C3 to C5), the

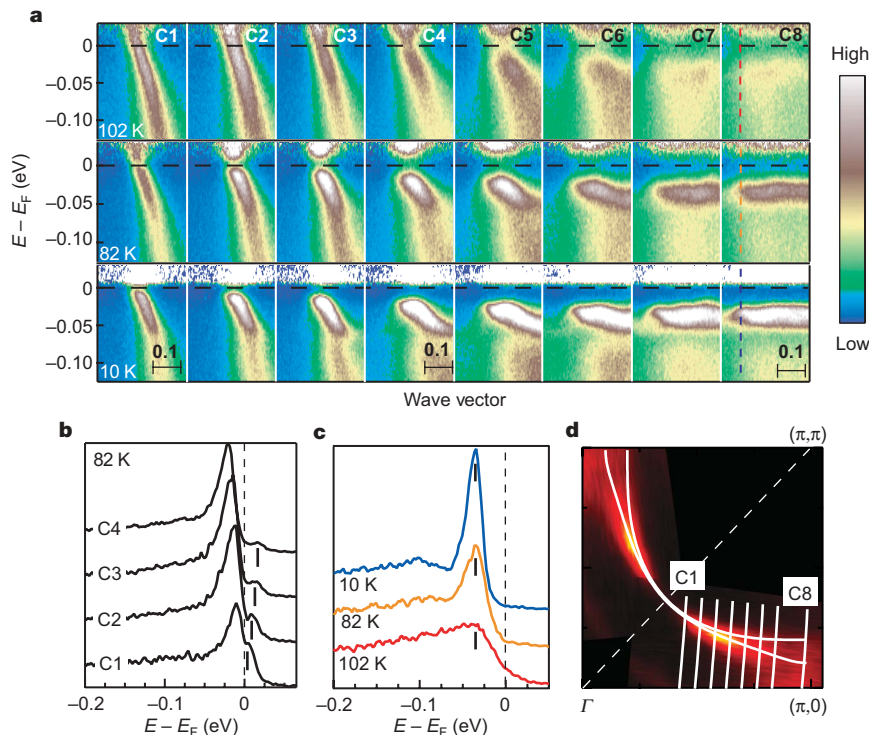
<sup>1</sup>Department of Physics, Applied Physics, and Stanford Synchrotron Radiation Laboratory, Stanford University, Stanford, California 94305, USA. <sup>2</sup>Advanced Light Source, Lawrence Berkeley National Lab, Berkeley, California 94720, USA. <sup>3</sup>Department of Applied Physics, University of Tokyo, Bunkyo-ku, Tokyo 113-8656, Japan. <sup>4</sup>Department of Physics, University of Waterloo, Ontario N2L 3G1, Canada.

temperature dependence of the gap across  $T_c$  becomes less pronounced, implying a smooth transition from one group to the other. With these two rather different temperature variations, a non-trivial temperature-dependent evolution of the gap function  $|\Delta_k(T)|$  along the Fermi surface can be sketched. As shown by the 82 K data in Fig. 3b, a gap consistent with a simple  $d_x^2 - y^2$  form,  $|\cos k_x - \cos k_y|/2$ , begins to develop near the node at a temperature just below  $T_c$ , whereas the gap near the antinode deviates from this nodal region  $d$ -wave gap. When the system is cooled well below  $T_c$ , the momentum dependence of the gap along the entire Fermi surface seems to be consistent with the simple  $d_x^2 - y^2$  form, at least for this doping. This non-trivial temperature evolution is another surprise associated with the discovery of the superconducting gap near the nodal region. We note that the value of  $2\Delta_{k=\text{antinode}}(T=10\text{ K})/k_B T_c$  (where  $k_B$  is the Boltzmann constant) of this  $d$ -wave gap is about 9, which is still much larger than the value  $\sim 4.12$  predicted by weak-coupling  $d$ -wave BCS theory.

In Fig. 3c, d, data for an underdoped sample with  $T_c = 75\text{ K}$  and for an overdoped sample with  $T_c = 86\text{ K}$  are shown. The temperature dependence of the gap function is consistent with that of the UD92K sample except that the gapless region above  $T_c$  extends with increasing doping. This change in the gap function in the superconducting state is qualitatively different from the simple mean field behaviour with a temperature-independent pairing interaction, in which the momentum dependence of the gap should not change at temperatures below  $T_c$ . Thus, the observed temperature-dependent evolution of the gap function implies an intriguing relation between

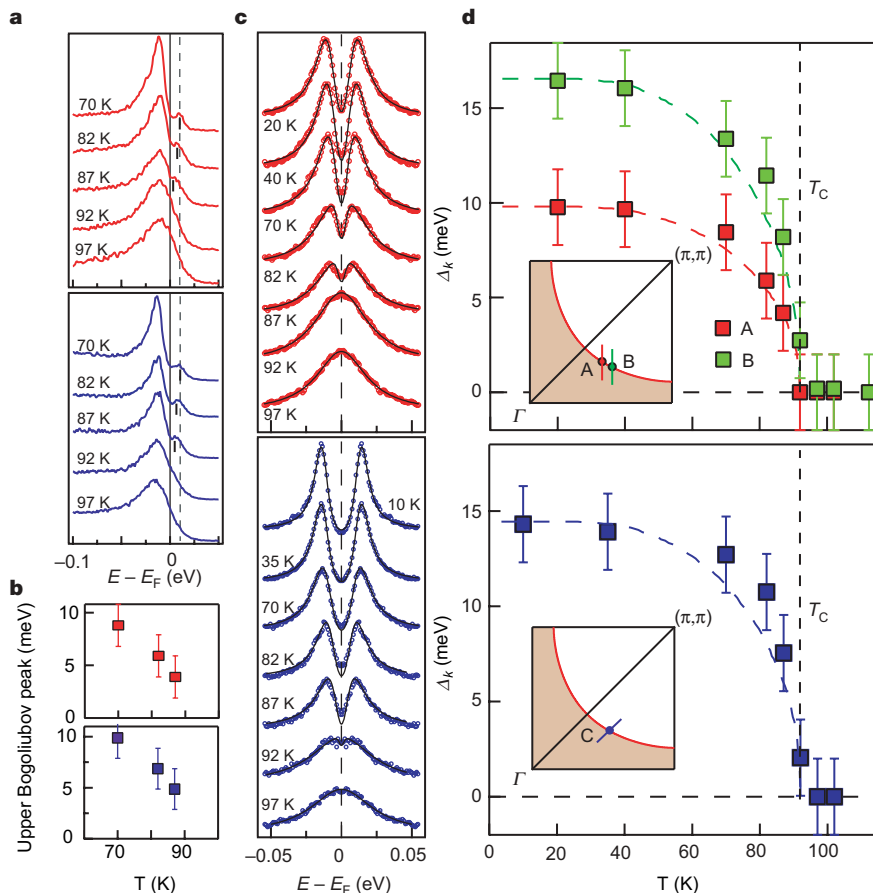
the superconducting gap and pseudogap. We note that in heavily underdoped samples, in which the pseudogap is much more pronounced than the superconducting gap, the gap function can only evolve into a U shape instead of a simple  $d$ -wave form at our lowest achievable temperature (see Fig. 3c and ref. 17). In Fig. 4 we summarize schematically the temperature-dependent evolution of the gap function in the three samples with different doping levels that we have studied.

It seems impossible to explain our data by a single gap. We are not aware of any mechanism that would create an energy gap that opens at different temperatures on the same sheet of the Fermi surface. In addition, the temperature-dependent evolution of the gap function along the Fermi surface in the superconducting state also seems very difficult to reconcile within a single-gap model. It seems more reasonable to assume the existence of two energy gaps. The energy gap opening at  $T_c$  near the nodal region is associated with the order parameter of the superconducting state, whereas the pseudogap near the antinodal region represents an energy scale associated with a different mechanism that may or may not be related to superconductivity. This is consistent with the two-gap picture derived from our recent doping dependence measurements on heavily underdoped Bi2212 (ref. 17). In addition, several other spectroscopy experiments, such as Andreev reflection<sup>11</sup>, intrinsic tunnelling spectroscopy<sup>13</sup> and femtosecond spectroscopy<sup>14</sup>, were also interpreted as a gap opening at  $T_c$ , which was difficult to understand in the context of previous ARPES<sup>19–21</sup> and scanning tunnelling microscope<sup>22</sup> results dominated by the antinodal region (see



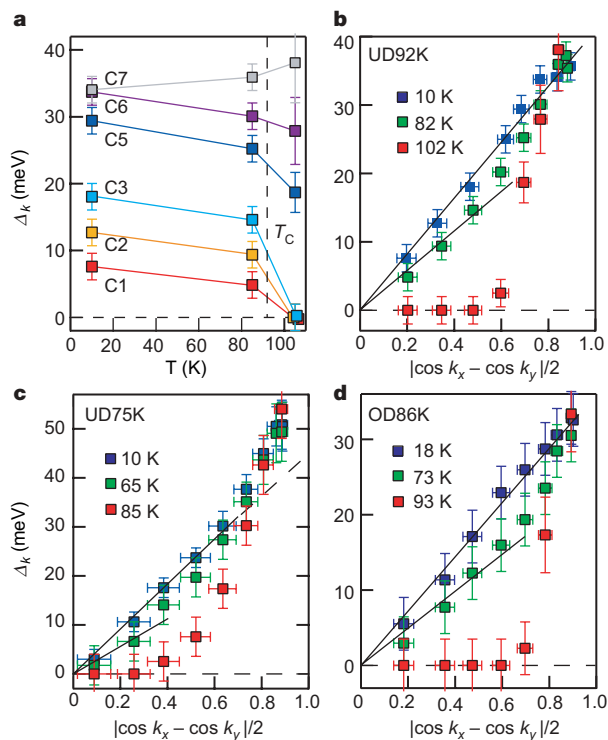
**Figure 1 | Temperature and momentum dependence of the low energy excitations in slightly underdoped Bi2212 ( $T_c = 92\text{ K}$ ).** **a**, Image plots of the Fermi–Dirac-function divided ARPES spectrum (see Supplementary Information) along the Fermi surface taken at three different temperatures: above  $T_c$  (102 K; top panels), right below  $T_c$  (82 K; middle panels), and well below  $T_c$  (10 K; bottom panels). The high-intensity region represents the band dispersion along the cutting directions as indicated by solid white lines in **d**. In the Fermi–Dirac-function divided spectrum, this high-intensity region either breaks at  $E_F$  if there is an energy gap or passes  $E_F$  if there is no detectable gap. Scale bar,  $0.1\pi$  (in units of lattice constant). **b**, Raw EDCs near the Fermi crossing point ( $k_F$ ) for C1 to C4 at 82 K. The short vertical lines indicate the thermally populated Bogoliubov band above  $E_F$ . The

Bogoliubov band dispersion is a signature of the superconducting state. **c**, Temperature dependence of raw EDCs near the  $k_F$  of C8. Short vertical lines indicate the gap energy. The momentum position of EDCs is indicated by the dashed lines in the C8 panels in **a**. A sharp peak in the spectrum can be observed just below  $T_c$ , although the gap size remains about the same across  $T_c$ . **d**, A partial Fermi-surface mapping measured at 102 K in a quadrant of the first Brillouin zone. The map is obtained by integrating raw spectra over the energy window  $E_F \pm 10\text{ meV}$  and symmetrizing with respect to the diagonal of the Brillouin zone. The intensity near the nodal direction (white dashed line) is suppressed because of the matrix element effect under this experimental setup.



**Figure 2 | Detailed temperature dependence of the superconducting gap near the nodal region of underdoped Bi2212 ( $T_c = 92$  K) measured under two different experimental configurations.** Data shown in each upper panel were measured along cuts parallel to the  $(\pi, 0) - (\pi, \pi)$  direction, using 22.7-eV photons and an energy resolution of 5 meV. The data in each lower panel were measured along cuts parallel to the  $(0, 0) - (\pi, \pi)$  direction, using 7-eV photons and an energy resolution of 3.2 meV.

**a**, Temperature dependence of raw EDCs near the Fermi-surface crossing points A (upper panel) and C (lower panel). The short bars indicate the peak of the thermally populated upper Bogoliubov band, whose position at 70 K is marked by the dashed line as a reference. When the temperature approaches  $T_c$ , the peak of the thermally populated Bogoliubov band moves towards  $E_F$  and then disappears above  $T_c$ . **b**, Temperature dependence of the peak position of the thermally populated upper Bogoliubov band in the raw spectra shown in **a**, which extrapolates to zero near  $T_c$ . The behaviours shown in **a** and **b** suggest that the magnitude of the superconducting gap decreases when the temperature approaches  $T_c$  and vanishes above  $T_c$ . **c**, Temperature dependence of the symmetrized EDCs at Fermi-surface crossing points A (upper panel) and C (lower panel) superimposed on their fit (black curves) to the phenomenological model<sup>23</sup>. At temperatures above  $T_c$ , the two peaks of the symmetrized EDCs merge into one peak, suggesting the collapse of the gap. **d**, Temperature dependence of the fitted gap size at Fermi-surface crossing points near the node, as indicated in the insets. The dashed lines show the temperature dependence of the superconducting gap based on weak-coupling BCS theory and serve as a guide to the eye for our data. The  $\Delta_k(T = 0)$  values for the BCS curves are adjusted independently so as to fit the data at different locations. The error bars are estimated to be  $\pm 2$  meV accounting for the uncertainties from the fitting procedure ( $\pm 0.5$  meV), the Fermi energy calibration ( $\pm 0.5$  meV) and an additional 100% margin.

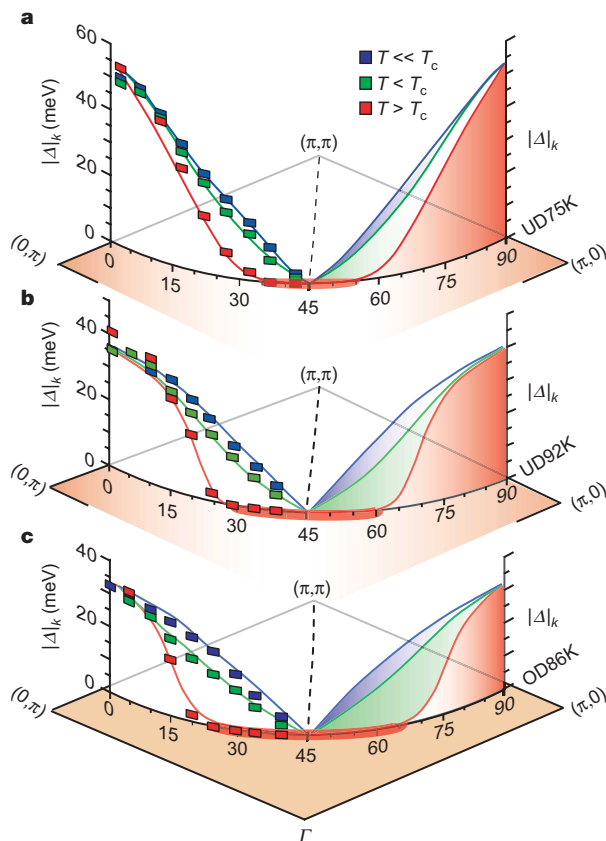


**Figure 3 | Temperature dependence of the gap profile.** **a**, The fitted gap values from the underdoped sample with  $T_c = 92$  K, for selected Fermi surface crossings as defined in Fig. 1d. The discrete temperature points should not be regarded as a gap closing at 102 K. **b**, Plot of the fitted gap values against the simple  $d_{x^2 - y^2}$  functional form,  $|\cos k_x - \cos k_y|/2$ , from the underdoped sample with  $T_c = 92$  K. The lines are guides to the eye indicating the expected momentum dependence of a simple  $d_{x^2 - y^2}$  form. **c**, **d**, The temperature-dependent evolution of the gap function for an underdoped sample with  $T_c = 75$  K (**c**) and an overdoped sample with  $T_c = 86$  K (**d**). The error bars for the superconducting gap are the same as those described in the legend to Fig. 2, whereas those of the pseudogap are larger because of the larger uncertainty from the fitting procedure.

Supplementary Information). We argue that these experimental probes are more sensitive to the nodal region, where the gap opens at  $T_c$ , as demonstrated in our data. Notably, the temperature dependence of the scanning tunnelling microscope spectrum has recently been revisited and the coexistence of two energy gaps in underdoped cuprates has been suggested<sup>15,16</sup>. In particular, a normalization procedure reveals that one of the gaps disappears at  $T_c$  (ref. 16), which is consistent with our finding near the nodal region.

What is the relationship between the observed two gaps? On the one hand, the distinct temperature and doping dependence seem to suggest a competing nature between the nodal BCS-like superconducting gap and the antinodal pseudogap; on the other hand, the evolution of the gap profile into a simple  $d$ -wave form at low temperature for UD92K and OD86K samples seems to suggest an intimate relationship between the superconducting gap and the





**Figure 4 | Schematic illustrations of the gap function evolution for three different doping levels.** **a**, Underdoped sample with  $T_c = 75$  K. **b**, Underdoped sample with  $T_c = 92$  K. **c**, Overdoped sample with  $T_c = 86$  K. At 10 K above  $T_c$  there exists a gapless Fermi arc region near the node; a pseudogap has already fully developed near the antinodal region (red curves). With increasing doping, this gapless Fermi arc elongates (thick red curve on the Fermi surface), as the pseudogap effect weakens. At  $T < T_c$  a  $d$ -wave like superconducting gap begins to open near the nodal region (green curves); however, the gap profile in the antinodal region deviates from the simple  $d_x^2 - y^2$  form. At a temperature well below  $T_c$  ( $T \ll T_c$ ) the superconducting gap with the simple  $d_x^2 - y^2$  form eventually extends across entire Fermi surface (blue curves) in **b** and **c** but not in **a**.

pseudogap. Theoretical calculations, in which the pseudogap is ascribed to a charge density wave competing with the superconducting state<sup>6,7</sup>, demonstrate a similar temperature dependence and doping dependence of the gap profile shown here and in our recent study on a heavily underdoped system<sup>17</sup>. Similarly, theories that treat a pseudogap as preformed Cooper pairs could also predict a two-gap behaviour. For example, a bipolaron theory<sup>28</sup> has shown two energy gaps with distinct temperature dependence: one gap opens at  $T_c$  in a BCS fashion and the other is essentially temperature independent across  $T_c$ . However, these theories would require the two gaps to add in quadrature, a feature inconsistent with the antinodal data, at least for the UD92K and OD86K samples. The other theories<sup>3–5</sup> have not yet developed detailed temperature dependence for a direct comparison with this experiment. In all cases, our finding would put a strong constraint on theory.

Received 28 May; accepted 4 September 2007.

1. Timusk, T. & Statt, B. The pseudogap in high-temperature superconductors: an experimental survey. *Rep. Prog. Phys.* **62**, 61–122 (1999).

2. Damascelli, A., Hussain, Z. & Shen, Z. X. Angle-resolved photoemission of cuprate superconductors. *Rev. Mod. Phys.* **75**, 473–541 (2003).
3. Emery, V. J. & Kivelson, S. A. Importance of phase fluctuations in superconductors with small superfluid density. *Nature* **374**, 434–437 (1995).
4. Wen, X. G. & Lee, P. A. Theory of underdoped cuprates. *Phys. Rev. Lett.* **76**, 503–506 (1996).
5. Chakravarty, S., Laughlin, R. B., Dirk, K., Morr, D. K. & Nayak, C. Hidden order in the cuprates. *Phys. Rev. B* **63**, 094503 (2001).
6. Benfatto, L., Caprara, S. & DiCastro, C. Gap and pseudogap evolution within the charge-ordering scenario for superconducting cuprates. *Eur. Phys. J. B* **17**, 95–102 (2000).
7. Li, J.-X., Wu, C.-Q. & Lee, D.-H. Checkerboard charge density wave and pseudogap of high- $T_c$  cuprate. *Phys. Rev. B* **74**, 184515 (2006).
8. Millis, A. J. Gaps and our understanding. *Science* **314**, 1888–1889 (2006).
9. Opel, M. et al. Carrier relaxation, pseudogap, and superconducting gap in high- $T_c$  cuprates: A Raman scattering study. *Phys. Rev. B* **61**, 9752–9774 (2000).
10. Le Tacon, M. et al. Two energy scales and two distinct quasiparticle dynamics in the superconducting state of underdoped cuprates. *Nature Phys.* **2**, 537–543 (2006).
11. Deutscher, G. Coherence and single-particle excitations in the high-temperature superconductors. *Nature* **397**, 410–412 (1999).
12. Svistunov, V. M., Tarenkov, V., Yu., D'Yachenko, A. I. & Hattar, E. Temperature dependence of the energy gap in Bi2223 metal oxide superconductor. *JETP Lett.* **71**, 289–292 (2000).
13. Krasnov, V. M., Yurgens, A., Winkler, D., Delsing, P. & Claeson, T. Evidence of coexistence of the superconducting gap and pseudogap in Bi-2212 from intrinsic tunnelling spectroscopy. *Phys. Rev. Lett.* **84**, 5860–5863 (2000).
14. Demsar, J., Hudej, R., Karpinski, J., Kabanov, V. V. & Mihailovic, D. Quasiparticle dynamics and gap structure in  $\text{HgBa}_2\text{CaCu}_2\text{O}_{8+\delta}$  investigated with femtosecond spectroscopy. *Phys. Rev. B* **63**, 054519 (2001).
15. Gomes, K. K. et al. Visualizing pair formation on the atomic scale in the high- $T_c$  superconductor  $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ . *Nature* **447**, 569–572 (2007).
16. Boyer, M. C. et al. Imaging the two gaps of the high-temperature superconductor  $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ . *Nature Phys.* doi:10.1038/nphys725 (in the press).
17. Tanaka, K. et al. Distinct Fermi-momentum-dependent energy gaps in deeply underdoped Bi2212. *Science* **314**, 1910–1913 (2006).
18. Kondo, T. et al. Evidence for two energy scales in the superconducting state of optimally doped  $(\text{Bi,Pb})_2(\text{Sr,La})_2\text{CuO}_{6+\delta}$ . *Phys. Rev. Lett.* **98**, 267004 (2007).
19. Loeser, A. G. et al. Excitation gap in the normal state of underdoped  $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ . *Science* **273**, 325–329 (1996).
20. Norman, M. R. et al. Destruction of the Fermi surface in underdoped high- $T_c$  superconductors. *Nature* **392**, 157–160 (1998).
21. Loeser, A. G. et al. Temperature and doping dependence of the Bi-Sr-Ca-Cu-O electronic structure and fluctuation effects. *Phys. Rev. B* **56**, 14185–14189 (1997).
22. Renner, Ch., Revaz, B., Genoud, J.-Y., Kadowaki, K. & Fischer, O. Pseudogap precursor of the superconducting gap in under- and overdoped  $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ . *Phys. Rev. Lett.* **80**, 149–152 (1998).
23. Norman, M. R., Randeria, M., Ding, H. & Campuzano, J. C. Phenomenology of the low-energy spectral function in high- $T_c$  superconductors. *Phys. Rev. B* **57**, R11093–R11096 (1998).
24. Hosseini, A. Microwave spectroscopy of thermally excited quasiparticle in  $\text{YBa}_2\text{Cu}_3\text{O}_{6.99}$ . *Phys. Rev. B* **60**, 1349–1359 (1999).
25. Krishana, K., Harris, J. M. & Ong, N. P. Quasiparticle mean free path in  $\text{YBa}_2\text{Cu}_3\text{O}_7$  measured by the thermal Hall conductivity. *Phys. Rev. Lett.* **75**, 3529–3532 (1995).
26. Valla, T. et al. Fine details of the nodal electronic excitations in  $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ . *Phys. Rev. B* **73**, 184518 (2006).
27. Yamasaki, T. et al. Unmasking the nodal quasiparticle dynamics in cuprate superconductors using low-energy photoemission. *Phys. Rev. B* **75**, 140513 (2007).
28. Alexandrov, A. S. & Andreev, A. F. Gap and subgap tunneling in cuprates. Preprint at arXiv:cond-mat/0005315v3 [cond-mat.supr-con].

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank R. Moore for experimental assistance, and D. J. Scalapino, S. Kivelson and T. K. Lee for helpful discussions. This work is supported by the DOE Office of Basic Energy Science, Division of Materials Science and Engineering, and the National Science Foundation. ARPES experiments were performed at the Stanford Synchrotron Radiation Laboratory (SSRL), which is operated by the Department of Energy Office of Basic Energy Science.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to Z.X.S. (zxshen@stanford.edu) or W.S.L. (leews@stanford.edu).

# Radio-frequency scanning tunnelling microscopy

U. Kemiktarak<sup>1</sup>, T. Ndukum<sup>3</sup>, K. C. Schwab<sup>3</sup> & K. L. Ekinci<sup>2</sup>

The scanning tunnelling microscope (STM)<sup>1</sup> relies on localized electron tunnelling between a sharp probe tip and a conducting sample to attain atomic-scale spatial resolution. In the 25-year period since its invention, the STM has helped uncover a wealth of phenomena in diverse physical systems—ranging from semiconductors<sup>2,3</sup> to superconductors<sup>4</sup> to atomic and molecular nanosystems<sup>5–9</sup>. A severe limitation in scanning tunnelling microscopy is the low temporal resolution, originating from the diminished high-frequency response of the tunnel current readout circuitry. Here we overcome this limitation by measuring the reflection from a resonant inductor–capacitor circuit in which the tunnel junction is embedded, and demonstrate electronic bandwidths as high as 10 MHz. This ~100-fold bandwidth improvement on the state of the art translates into fast surface topography as well as delicate measurements in mesoscopic electronics and mechanics. Broadband noise measurements across the tunnel junction using this radio-frequency STM have allowed us to perform thermometry at the nanometre scale. Furthermore, we have detected high-frequency mechanical motion with a sensitivity approaching ~15 fm Hz<sup>−1/2</sup>. This sensitivity is on par with the highest available from nanoscale optical and electrical displacement detection techniques, and the radio-frequency STM is expected to be capable of quantum-limited position measurements.

In typical STM operation, the tip is scanned over the sample surface, while keeping the tunnel resistance  $R_T$  constant in the range  $1\text{ M}\Omega \leq R_T \leq 1\text{ G}\Omega$ . The stray capacitance  $C_p$  in the macroscopic wiring limits the bandwidth, in most cases, to  $1/2\pi R_T C_p \approx 10\text{ kHz}$ . This is far from the fundamental limit,  $I_T/e$ , which is the tunnelling rate of a single electron determined by its charge  $e$ . For a typical tunnel current  $I_T \approx 1\text{ nA}$ , this suggests an available bandwidth of ~1 GHz. So far, researchers have demonstrated several avenues towards achieving better time resolution with the STM. Direct radio-frequency (RF) measurements on the tunnel current have shown signatures of electron spin resonance at high frequencies<sup>10,11</sup>. Using several amplifier stages instead of a single trans-impedance amplifier<sup>12,13</sup> improves the time resolution by roughly an order of magnitude, but with an undesirable reduction in the signal-to-noise ratio. Optically based pump-probe techniques<sup>14,15</sup>, in rather complex experimental geometries, have allowed for picosecond time resolution in the tunnel current after long averaging times. Our approach here is to read out the tunnel junction resistance directly by engineering an inductor–capacitor (LC) impedance transformation network. Such impedance matching has been used to dramatically improve the read-out bandwidth and coupling efficiency of high impedance nanoscale devices such as single electron transistors<sup>16</sup>, atomic point contacts<sup>17</sup> and nanomechanical resonators<sup>18</sup>.

The essential components of the RF-STM are shown in the block diagram of Fig. 1a. The large tunnel junction resistance  $R_T$  is transformed down by an LC impedance transformation network, also called the tank circuit, which is coupled to a 50- $\Omega$  low-noise RF amplifier. In our implementation, the RF circuit is placed on a small

circuit board and mounted in a variable temperature STM; it does not disturb the conventional low-frequency circuit of the STM (see Methods for a description of the apparatus). At the resonance frequency  $f_{LC}$  of the LC circuit, the total impedance becomes  $Z_{RLC} = Z_{LC}^2/R_T$ , where the inductance  $L$  and capacitance  $C$  determine  $Z_{LC}$ :  $Z_{LC} = \sqrt{L/C}$ . By engineering  $Z_{RLC}$  appropriately, one can couple the small changes in  $R_T$  efficiently into the high-frequency measurement circuit.

Figure 1b illustrates the underlying principle of RF-STM operation. The inset shows the reflected power from the resonant LC circuit at different  $R_T$  as a function of frequency, as the tip is lowered towards a flat region on a Au surface. The changes in  $R_T$  result in changes in the damping of the resonant LC circuit (made up of a 2  $\mu\text{H}$  chip-inductor and a 0.33 pF chip-capacitor), thereby changing the reflected power. The tip–sample capacitance is ~1 fF (ref. 19) and does not load the circuit. With additional stray capacitances on the circuit board, the tank circuit resonates at  $f_{LC} \approx 115\text{ MHz}$ . The main panel of Fig. 1b shows the on-resonance amplitude of the reflection coefficient,  $|I(f_{LC})|$ , as a function of  $R_T$ ; here  $I(f_{LC}) = [Z_{RLC}(f_{LC}) - 50\Omega]/[Z_{RLC}(f_{LC}) + 50\Omega]$ . In RF-STM operation, the displayed change in  $|I(f_{LC})|$  with  $R_T$ , namely, with the tunnel gap  $z$ , provides transduction of displacements into electrical signals. An approximate  $|I(f_{LC})|$  versus  $z$  calibration can be naively obtained (Fig. 1b, upper axis) by assuming point contact at  $R_0 \approx 25\text{ k}\Omega$  and an exponentially increasing tunnel resistance,  $R_T \approx R_0 e^{2\kappa z}$ . For the Au surface,  $\kappa \approx 1\text{ \AA}^{-1}$  is obtained from further measurements (see below, and Supplementary Information section 2).

The most straightforward way to demonstrate fast imaging by the RF-STM is in the ‘constant height’ mode. Here, after reducing the STM feedback gain, RF power at frequency  $f_{LC}$  is launched towards the tip at each scan grid point; the reflected power is then measured (see Methods). Owing to the reduced feedback, surface height variations and the subsequent  $R_T$  variations result in contrasts in the reflected power image. Ideally, the minimum measurement time at each scan point is set by the 2-MHz bandwidth of the tank circuit (half-width at half-maximum, HWHM), and is ~0.1  $\mu\text{s}$ . In practice, our measurement speed is limited by the kHz-range resonances of the STM tube scanner (see Methods for further discussion of bandwidth). Figure 1c and d show images of a 100 nm  $\times$  100 nm area of a Au-on-mica sample taken by constant-current STM and RF-STM, respectively. The ~2.3- $\text{\AA}$  step heights observed are close to those of single Au(111) steps<sup>20</sup>. The RF-STM image has negative contrast as compared to the constant-current STM image: the higher the feature on the surface, the smaller the  $R_T$  and hence, the smaller the reflected power.

To calibrate the RF circuit, we have used the shot noise in the tunnel current. In shot noise measurements, we remove the directional coupler and analyse the voltage fluctuations with a spectrum analyser (SA; Fig. 2a inset). The main panel of Fig. 2a shows a typical measurement of the shot noise power collected in the tank circuit. The first step in the calibration is the extraction of the tank circuit  $Q$ .

<sup>1</sup>Department of Physics, <sup>2</sup>Department of Aerospace and Mechanical Engineering, Boston University, Boston, Massachusetts 02215, USA. <sup>3</sup>Department of Physics, Cornell University, Ithaca, New York 14853, USA.

The spectral density  $S_I$  of the shot noise satisfies<sup>21</sup>

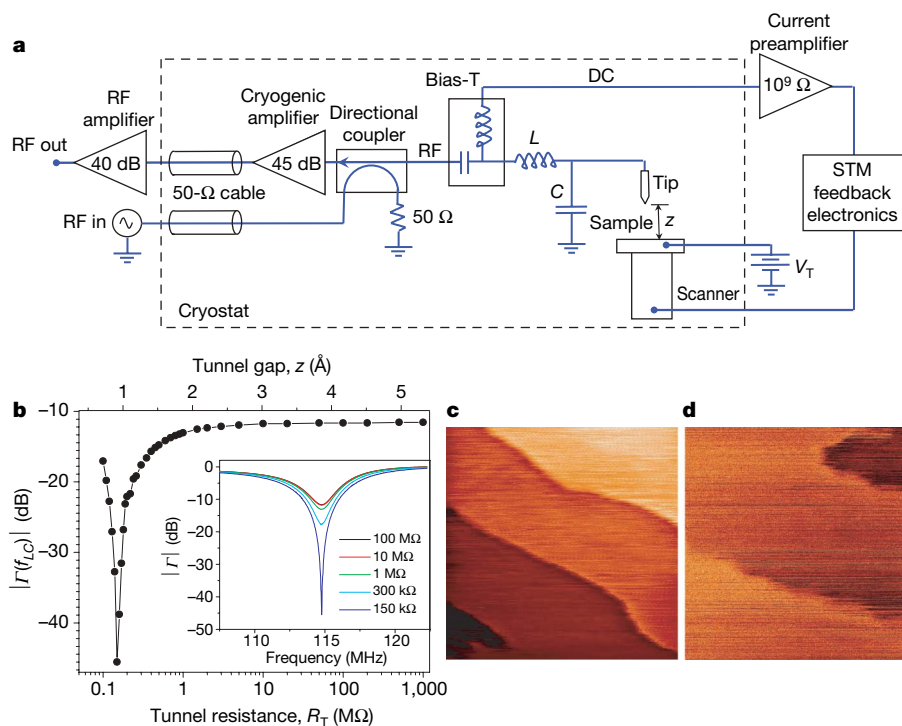
$$S_I = 2eI_T \coth \frac{eI_T R_T}{2k_B T} \quad (1)$$

This expression depends upon  $R_T$ ,  $I_T$  and the absolute temperature  $T$  as well as the Boltzmann constant  $k_B$  and the electronic charge  $e$ . Measurement of  $S_I$  can be used to determine<sup>22</sup>: (1) the gain-bandwidth product (GBW) of the amplifier when  $I_T R_T > k_B T/e$ ; and (2) the noise injected by the circuit when  $I_T R_T \approx 0$ . Figure 2b demonstrates the calibration process carried out at two different  $R_T$  values at 77 K. After accounting for the excess noise and the GBW, the data are fitted by the function in equation (1) using  $T = 77$  K.

Beyond calibration, shot noise in the tunnel current can be used for primary thermometry, where the extracted temperature is only a function of fundamental physical constants with no need for further calibration. Shot-noise-based primary thermometry was recently demonstrated in lithographic metallic tunnel junctions<sup>22</sup>. Here we extend this approach to nanometre-scale tunnel junctions using the RF-STM. In these experiments, we keep the tip-sample gap fixed and measure the shot noise power at three different bath temperatures, to which both the tip and the sample are anchored. Figure 2c displays the effect of temperature on the noise power in RF-STM measurements. After removing the GBW and the excess noise, the temperature can be extracted from fits to equation (1). The expected

temperature dependence in Fig. 2c encourages us to propose thermal imaging based on an RF-STM. In such an endeavour, one could measure the shot noise as a function of position over a sample surface, where there are local temperature deviations from the bath temperature. A straightforward analysis (see Supplementary Information section 1 and Supplementary Fig. 1) shows that the measured 'effective' shot noise temperature becomes  $T_{\text{eff}} \approx (T_T + T_S)/2$  when the tip and sample are at different respective temperatures  $T_T$  and  $T_S$ —thereby conveying information about the local temperature variations. We believe that the RF-STM, with its broad bandwidth and nanometre-scale spatial resolution, is particularly suited for rapid thermal imaging<sup>23</sup> of technologically important devices—for example, advanced semiconductor chips, where heat flow at the nanoscale is currently the limiting factor on density and clock speed. Our estimates suggest that, at 300 K, with a background noise of 70 K, a 100 pixel  $\times$  100 pixel thermal image with  $\sim 1$  K temperature resolution should be achievable in  $\sim 100$  s (see Methods and Supplementary Information section 1). The spatial resolution in such a measurement should be determined by the nanoscale tunnel junction area.

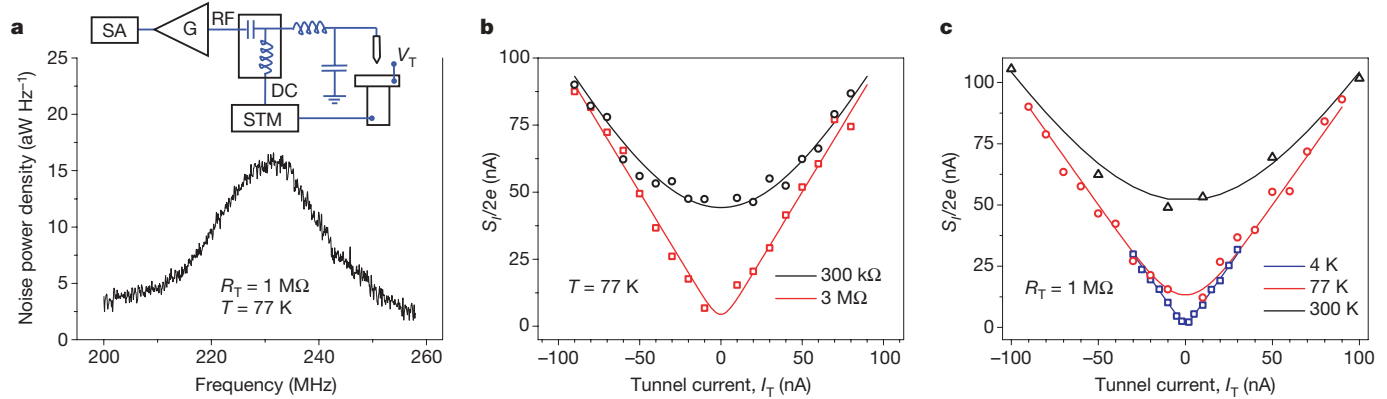
The RF-STM can be used as a local broadband displacement sensor with ultrahigh displacement sensitivity<sup>24</sup>—in theory, limited by quantum mechanics<sup>25–27</sup>. In order to demonstrate this, we have performed displacement measurements on two different systems: a flat Au sample shaken by a piezoelectric actuator, and a micromechanical



**Figure 1 | Diagram and operation of the RF-STM.** **a**, The tunnel junction is embedded in an LC tank circuit. Low-frequency tunnel current is separated from the RF signal and fed into the STM feedback circuit to control the tip-sample separation  $z$ . **b**, Reflection coefficient  $\Gamma$  as a function of tunnel junction resistance  $R_T$  on flame annealed Au deposited on mica. The inset shows  $\Gamma$  as a function of frequency at different  $R_T$ . At large  $R_T$ , the losses in the tank dominate the response; most of the incoming power is dissipated and the rest reflected. As  $R_T$  is reduced by bringing the tip closer to the surface, a better match to 50  $\Omega$  is obtained and the tank resonance-dip gets deeper, indicating more efficient RF power coupling to the tunnel junction. At  $R_T \approx 150$  k $\Omega$ , best matching is achieved. Decreasing  $R_T$  further increases the impedance mismatch. Changes in  $\Gamma$  with  $R_T$  at the tank circuit resonance frequency,  $f_{LC} \approx 115$  MHz (main panel), enable sensitive transduction of  $z$  into electrical signals. Starting from point contact ( $R_0 \approx 25$  k $\Omega$ ) and assuming an exponential decay  $R_T \approx R_0 e^{2\kappa z}$  with  $\kappa \approx 1 \text{ \AA}^{-1}$ , a  $z$  calibration (top x-axis) is obtained. On this surface, the  $\kappa \approx 1 \text{ \AA}^{-1}$  value and the exponential

dependence are verified experimentally for  $R_T = 1, 2$  and 4 M $\Omega$  (see Supplementary Information section 2). The available bandwidth with this particular LC tank is  $\sim 2$  MHz (HWHM). **c**, Constant-current STM image of 100 nm  $\times$  100 nm Au surface, showing several steps with height  $\sim 2.3$   $\text{\AA}$ . Here,  $R_T \approx 3$  M $\Omega$ ,  $I_T = 2$  nA and  $V_T = 6$  mV. The scan speed is 1 s per line. A background was subtracted. **d**, RF-STM image of the same sample (at a different location) with  $R_T \approx 4$  M $\Omega$ ,  $I_T = 2$  nA and  $V_T = 8$  mV, and scan speed of 0.01 s per line. The scan area is again 100 nm  $\times$  100 nm. The displayed image is raw data. Because of the drifts and hysteresis in the scanner at high speeds, the exact same region shown in **c** could not be imaged. The RF-STM image has inverted contrast in comparison to the topography image. During imaging, the reduced feedback allows contrast to emerge at the step edges; then, feedback slowly attains the set d.c. tunnel current level on the terraces, reducing the contrast. The jump from a higher terrace to a lower one corresponding to an increase of  $\sim 2.3$   $\text{\AA}$  results in a  $\sim 0.5$  dB dip in the reflected power at  $R_T \approx 4$  M $\Omega$ . All the data in this figure were acquired at room temperature.





**Figure 2 | Primary noise thermometry using the RF-STM.** **a**, Typical RF noise power density at 77 K with  $R_T \approx 1 \text{ M}\Omega$  and  $V_T = 200 \text{ mV}$ . Here, a different circuit made up of a  $0.5 \mu\text{H}$  chip-inductor and a  $0.5 \text{ pF}$  chip-capacitor ( $f_{LC} \approx 230 \text{ MHz}$ ) enables a higher bandwidth of  $\sim 10 \text{ MHz}$  (HWHM). The inset shows the simplified measurement schematic (SA, spectrum analyser; G, low-noise RF amplifier). **b**, Normalized shot noise

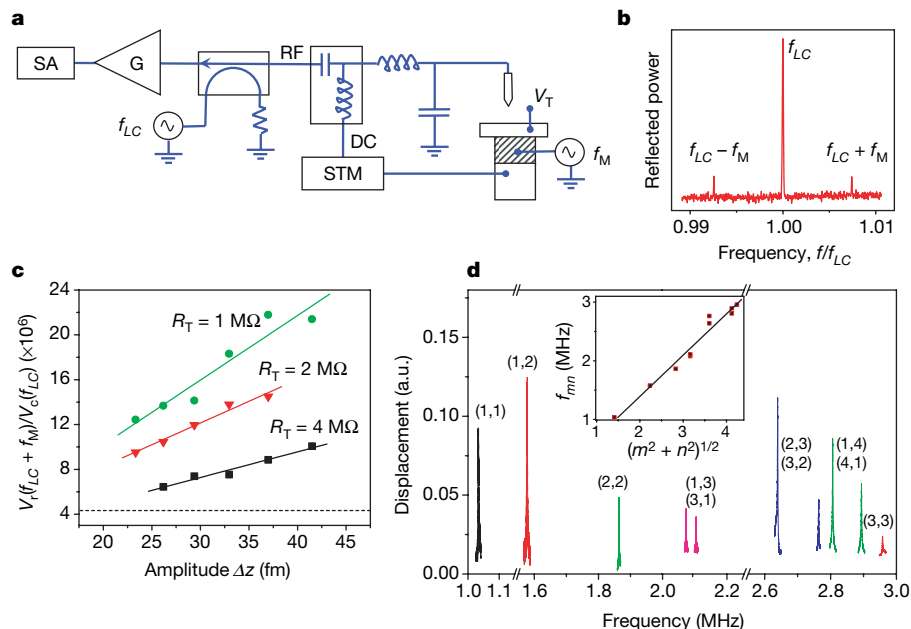
power spectral density  $S_I/2e$  at the tunnel junction displayed as a function of  $I_T$  at  $R_T = 300 \text{ k}\Omega$  and  $3 \text{ M}\Omega$ . At the measurement temperature of 77 K, both measurements give an excess noise of  $\sim 35 \text{ K}$ ; the gain-bandwidth (GBW) product of the amplifier is  $\sim 10^{11} \text{ Hz}$ . **c**,  $S_I/2e$  as a function of  $I_T$  at three different temperatures (300 K, 77 K and 4 K), but a fixed  $R_T \approx 1 \text{ M}\Omega$ . The continuous lines in **b** and **c** are fits to equation (1).

membrane excited at its resonant modes. Figure 3a and b respectively depict the simplified measurement set-up and a typical measurement obtained from this set-up. In the measurements, RF power at frequency  $f_{LC}$  is launched towards the tunnel junction, while the surface motion at frequency  $f_M$  modulates the reflected power. Data plotted in Fig. 3c are obtained from a collection of such reflection measurements on the Au surface as the surface oscillation amplitude  $\Delta z$  is increased. For small  $\Delta z$ , the sideband peak amplitude depends linearly on  $\Delta z$ . It is possible to obtain an estimate for the displacement sensitivity  $\sqrt{S_z(f_M)}$  achievable by the RF-STM from this measurement. The spectral density  $S_{V_r}(f_{LC} \pm f_M)$  of the RF voltage noise in

the sideband limits the displacement sensitivity as

$$\sqrt{S_{V_r}(f_{LC} \pm f_M)} \approx \frac{V_c}{2} \left. \frac{\partial |I|}{\partial z} \right|_{R_T} \sqrt{S_z(f_M)} \quad (2)$$

Here,  $V_c$  is the amplitude of the RF carrier at frequency  $f_{LC}$  and  $\partial|I|/\partial z|_{R_T}$  represents the change in  $I$  with  $z$  around the bias point. Limited by the amplifier noise floor, we estimate  $\sqrt{S_z} \approx 15 \text{ fm Hz}^{-1/2}$  when  $R_T \approx 1 \text{ M}\Omega$ . The data in Fig. 3c in conjunction with the data in Fig. 1b also allow a direct measurement of the decay constant as  $\kappa \approx 1 \text{ \AA}^{-1}$  (see Supplementary Information section 2).



**Figure 3 | RF-STM based displacement detection.** **a**, The measurement set-up. **b**, A typical spectrum of the RF-STM signal when the tunnel gap  $z$  is modulated by the mechanical motion of the surface placed under the tip. The carrier signal is applied at frequency  $f_{LC}$  and the surface is moved at frequency  $f_M$  by a piezoelectric actuator (shaded), separate from the STM scanner. The circuit used in this experiment is identical to that in Fig. 1. Experiments are at room temperature. **c**, Displacement measurements on a flat Au surface. Each data point is the ratio of the single-sideband reflection amplitude  $V_r(f_{LC} + f_M)$  to the carrier amplitude  $V_c(f_{LC})$  plotted against the r.m.s. motion amplitude  $\Delta z$  at  $1 \text{ MHz}$ . A linear dependence of  $V_r(f_{LC} + f_M)$  on  $\Delta z$  is observed at  $R_T = 1 \text{ M}\Omega$ ,  $2 \text{ M}\Omega$  and  $4 \text{ M}\Omega$ , tunnel current  $I_T = 2 \text{ nA}$ ,

and  $V_c \approx 130 \mu\text{V}$ . Linear fits to the data (solid lines) at each  $R_T$  go through the origin. The horizontal dashed line indicates the  $\sim 0.5 \text{ nV Hz}^{-1/2}$  noise floor set by the amplifier. The displacement sensitivity is obtained from the slope of each line; the best value ( $15 \text{ fm Hz}^{-1/2}$ ) is achieved at the lowest  $R_T$ . **d**, Swept frequency RF-STM measurements of the resonances of a micromechanical membrane. Here, the piezoelectric actuator is driven by a network analyser. The RF-STM signal is mixed-down with the carrier signal and is input into the detection port of the network analyser. The inset shows the measured eigenfrequencies as a function of mode number  $(m, n)$  of a membrane under tension. Note that the mode corresponding to  $(n, m) = (2, 1)$  is missing in the spectrum.

In a set of swept frequency measurements, we obtain the eigenmode frequencies of a membrane resonator (Fig. 3d). The micromechanical resonator is a ( $w \times l \times t \approx 65 \mu\text{m} \times 70 \mu\text{m} \times 40 \text{nm}$ ) silicon nitride membrane patterned with alignment marks and coated with a 25-nm-thick Au film (see Supplementary Information and Supplementary Fig. 3). In these measurements, the membrane motion is excited by the piezoelectric actuator; the RF-STM tip is positioned on a desired point on the moving membrane and the reflected power is monitored as a function of the drive frequency. The mode frequencies  $f_{mn}$  agree with optical interferometry measurements and accurately fit those of a membrane under tension,  $f_{mn} \propto \sqrt{m^2 + n^2}$  (see Supplementary Information section 3)—as shown in Fig. 3d inset. Beyond displacement detection, this technique would be especially useful for imaging nanoscale motion. We also note that we observe back-action of the tip on the membrane motion (for example, due to electrostatic forces) resulting in perturbation of the observed mechanical parameters. Both mode imaging and back-action will be the subject of future work.

We expect the RF-STM to be applied to a number of important scientific and technological problems. The simplicity of the technique adds to its attractiveness: (1) relatively minor modifications to an existing STM are required; (2) the RF components are readily available; and (3) the RF circuit can work under a wide range of conditions, for example, in high vacuum, at low temperature, and in high magnetic field. From a microscopy point of view, a second-generation RF-STM, which could scan at speeds comparable to its augmented electronic bandwidth, would be the next step in development work<sup>28</sup>. A broadband STM controller circuit, which can feedback on the RF signal, would also be useful.

The RF-STM is well suited to the study of the dynamics of electronic phenomena with possibly atomic-scale resolution. An excellent example of this is the investigation of electron spin resonance. The tunnel current modulation at the Larmor precession frequency  $f_{\text{ESR}}$  of the electrons can be measured conveniently by a RF-STM. A direct way to detect narrow-band ultrahigh-frequency current modulations at  $f_{\text{ESR}}$ , even with the current  $\sim 200 \text{MHz}$  tank frequencies, is to launch RF signals to the tunnel junction at  $\sim f_{\text{ESR}} + f_{\text{LC}}$  and detect the modulation sideband at  $\sim f_{\text{LC}}$ . This approach is limited only by the typical tunnelling time for electrons. RF-STM-based shot noise measurements on exotic superconductors may enable the fast determination of quasi-particle nature<sup>29</sup>.

Finally, electron tunnelling is expected to enable quantum-mechanically-limited position detection<sup>25–27</sup>. Research in this direction has begun to uncover interesting physics—in particular, the nature of back-action in detecting nanomechanical motion<sup>17,30</sup>. The RF-STM has the potential to further such studies, given that it is possible to tune the interaction strength and the position of the tip on the moving structure.

## METHODS SUMMARY

The STM head design is a standard one, which comprises several piezoelectric elements. The head can be inserted into a continuous flow cryostat with a base temperature of 2.6 K. The STM-cryostat sits on an optical table for vibration isolation and can be pumped down to pressures below 1 torr. In the experiment, the tunnel bias voltage is connected to the sample. The LC network for transforming the tunnel junction resistance is attached to the STM tip; the rest of the circuit elements are placed on a small circuit board and are connected to the LC circuit embedding the tunnel junction. For imaging, the reflected power from the resonant circuit is detected using a homodyne power detector. In the shot noise measurements, the RF excitation arm of the circuit is removed and the homodyne detector is replaced with a spectrum analyser. In the displacement sensitivity measurements, the full circuit is used; the spectrum of the reflected signal is measured using a spectrum analyser. In the membrane mode measurements, swept frequency measurements are performed using a network analyser.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 19 April; accepted 7 September 2007.

1. Binnig, G., Rohrer, H., Gerber, C. & Weibel, E. Surface studies by scanning tunneling microscopy. *Phys. Rev. Lett.* **49**, 57–61 (1982).

2. Strosio, J. A., Feenstra, R. M. & Fein, A. P. Electronic structure of the Si(111)2 × 1 surface by scanning-tunneling microscopy. *Phys. Rev. Lett.* **57**, 2579–2582 (1986).
3. Wolkow, R. A. & Avouris, P. Atom-resolved surface chemistry using scanning tunneling microscopy. *Phys. Rev. Lett.* **60**, 1049–1052 (1988).
4. Pan, S. H. *et al.* Imaging the effects of individual zinc impurity atoms on superconductivity in Bi<sub>2</sub>Sr<sub>2</sub>CaCu<sub>2</sub>O<sub>8+δ</sub>. *Nature* **403**, 746–750 (2000).
5. Eigler, D. M. & Schweizer, E. K. Positioning single atoms with a scanning tunnelling microscope. *Nature* **344**, 524–526 (1990).
6. Whitman, L. J., Strosio, J. A., Dragoset, R. A. & Celotta, R. J. Manipulation of adsorbed atoms and creation of new structures on room-temperature surfaces with a scanning tunneling microscope. *Science* **251**, 1206–1210 (1991).
7. Manoharan, H. C., Lutz, C. P. & Eigler, D. M. Quantum mirages formed by coherent projection of electronic structure. *Nature* **403**, 512–515 (2000).
8. Hornbaker, D. J. *et al.* Mapping the one-dimensional electronic states of nanotube peapod structures. *Science* **295**, 828–831 (2002).
9. Madhavan, V., Chen, W., Jamneala, T., Crommie, M. F. & Wingreen, N. S. Tunneling into a single magnetic atom: spectroscopic evidence of the Kondo resonance. *Science* **280**, 567–569 (1998).
10. Manassen, Y., Hamers, R. J., Demuth, J. E. & Castellano, A. J. Jr. Direct observation of the precession of individual parametric spins on oxidized silicon surfaces. *Phys. Rev. Lett.* **62**, 2531–2534 (1989).
11. Durkan, C. & Welland, C. E. Electronic spin detection in molecules using scanning-tunneling-microscopy-assisted electron spin resonance. *Appl. Phys. Lett.* **80**, 458–460 (2002).
12. Mamin, H. J., Birk, H., Wimmer, P. & Rugar, D. High-speed scanning tunneling microscopy: Principles and applications. *J. Appl. Phys.* **75**, 161–168 (1994).
13. Rost, M. J. *et al.* Scanning probe microscopes go video rate and beyond. *Rev. Sci. Instrum.* **76**, 053710 (2005).
14. Nunes, G. & Freeman, M. R. Picosecond resolution in scanning tunneling microscopy. *Science* **262**, 1029–1032 (1993).
15. Weiss, S., Ogletree, D. F., Botkin, D., Salmeron, M. & Chemla, D. S. Ultrafast scanning probe microscopy. *Appl. Phys. Lett.* **63**, 2567–2569 (1993).
16. Schoelkopf, R. J., Wahlgren, P., Kozhevnikov, A. A., Delsing, P. & Prober, D. The radio-frequency single-electron transistor (RF-SET): A fast and ultrasensitive electrometer. *Science* **280**, 1238–1242 (1998).
17. Flowers-Jacobs, N. E., Schmidt, D. R. & Lehnert, K. W. Intrinsic noise properties of atomic point contact displacement detectors. *Phys. Rev. Lett.* **98**, 096804 (2007).
18. Truitt, P. A., Hertzberg, J. B., Huang, C. C., Ekinci, K. L. & Schwab, K. C. Efficient and sensitive capacitive readout of nanomechanical resonator arrays. *Nano Lett.* **7**, 120–126 (2007).
19. Kurokawa, S. & Sakai, A. Gap dependence of the tip-sample capacitance. *J. Appl. Phys.* **83**, 7416–7423 (1998).
20. Hallmark, V. M., Chiang, S., Rabolt, J. F., Swalen, J. D. & Wilson, R. J. Observation of atomic corrugation on Au(111) by scanning tunneling microscopy. *Phys. Rev. Lett.* **59**, 2879–2882 (1987).
21. Birk, H., de Jong, J. M. & Schönenberger, C. Shot-noise suppression in the single-electron tunneling regime. *Phys. Rev. Lett.* **75**, 1610–1613 (1995).
22. Spietz, L., Lehnert, K. W., Siddiqi, I. & Schoelkopf, R. J. Primary electronic thermometry using the shot noise of a tunnel junction. *Science* **300**, 1929–1932 (2003).
23. Majumdar, A., Carrejo, J. P. & Lai, J. Thermal imaging using the atomic force microscope. *Appl. Phys. Lett.* **62**, 2501–2503 (1993).
24. Binnig, G., Quate, C. F. & Gerber, Ch. Atomic force microscope. *Phys. Rev. Lett.* **56**, 930–933 (1986).
25. Yurke, B. & Kochanski, G. P. Momentum noise in vacuum tunneling transducers. *Phys. Rev. B* **41**, 8184–8194 (1990).
26. Presilla, C., Onofrio, R. & Bocko, M. F. Uncertainty-principle noise in vacuum-tunneling transducers. *Phys. Rev. B* **45**, 3735–3743 (1992).
27. Clerk, A. A. & Girvin, S. M. Shot noise of a tunnel junction displacement detector. *Phys. Rev. B* **70**, 121303 (2004).
28. Xu, Y., MacDonald, N. C. & Miller, S. A. Integrated micro-scanning tunneling microscope. *Appl. Phys. Lett.* **67**, 2305–2307 (1995).
29. Jehl, X., Sanquer, M., Calemczuk, R. & Mailly, D. Detection of doubled shot noise in short normal-metal/ superconductor junctions. *Nature* **405**, 50–53 (2000).
30. Naik, A. *et al.* Cooling a nanomechanical resonator with quantum back-action. *Nature* **443**, 193–197 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank D. M. Karabacak for help with optical interferometry and A. Vandelay for discussions. This work was supported by the National Science Foundation through the Division of Materials Research (IMR Programme), the Division of Civil, Mechanical and Manufacturing Innovation (MDSE Programme) and the Cornell Center for Materials Research.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to K.L.E. ([ekinci@bu.edu](mailto:ekinci@bu.edu)).

## METHODS

**Apparatus.** The STM head design is a standard one, which comprises several piezoelectric elements. The sample plate is attached to a central piezoelectric tube by a spring clip. The central tube is supported by three piezoelectric tubes and springs. The central tube can slide along the  $z$ -axis by 'slip-stick' motion resulting from the supporting piezos; moreover, the sample plate can be moved in the  $x$ - $y$  plane by motion of the central piezo, thus enabling relative tip-sample motion along all three axes. The head is attached to the end of a 40-cm-long rod; this allows the insertion of the head into a continuous flow cryostat with a base temperature of 2.6 K. The STM-cryostat sits on an optical table for vibration isolation. The system can be pumped down to pressures below 1 torr. At room temperature, the experiments are either performed in an inert gas atmosphere or in vacuum. In cryogenic operation, the STM head is cooled by cold helium vapour, which flows from a liquid helium reservoir to the sample space through capillaries. The rate of this flow, and hence the base temperature, is controlled by a needle valve and a rotary pump.

In this experiment, the sample is connected to the tunnel bias voltage, while all the relevant RF and low-frequency circuitry are connected to the STM tip. The STM tip, a cut or etched Pt-Ir wire, is fixed onto a printed circuit board. This  $\sim 2\text{ cm} \times 2\text{ cm}$  board connects the tip to the  $LC$  tank circuit, the bias-T (Mini Circuits TCBT-2R5G) and the directional coupler (Mini Circuits DBTC-9-4+) as shown in Fig. 1a. The bias-T separates the low frequency part of the tunnelling signal from the high frequency part, and channels these two into the relevant circuits. The low frequency arm of the bias-T is connected to a current preamplifier by a coaxial cable  $\sim 24$  inches long. At the high frequency arm, there is a directional coupler that properly directs the incoming and reflected signals. After the directional coupler, there is a PCB to SMA connector, to which the low noise cryogenic RF amplifier (Miteq AFS3-00100200-09-CR-4) is connected. At room temperature, this amplifier has a gain of 45 dB and background noise temperature of  $\sim 70$  K. At 4 K, the gain goes up to 48 dB and the noise temperature goes down to  $\sim 5$  K. In the measurements, we used a second stage amplifier (Mini Circuits ZKL-1R5) outside the cryostat with a gain of 40 dB.

**Imaging.** For imaging, the RF reflectometry circuit in Fig. 1a is used after the addition of a homodyne power detector. First, feedback is reduced to its minimum value; this causes the tip to respond very slowly to surface height changes. A carrier signal at the resonance frequency  $f_{LC}$  of the  $LC$  circuit is launched towards

the tunnel junction. As the STM tip scans over the surface, the surface corrugations result in changes in  $R_T$  due to the slow feedback; this, subsequently, changes the reflected power. The reflected RF power is mixed with the carrier at  $f_{LC}$  using a mixer (Mini Circuits ZP-3+). The baseband signal is digitized using an analogue-to-digital (A/D) converter and plotted on the scan grid points on a computer. In RF-STM imaging of the Au surface, the tank circuit has a bandwidth (HWHM) of  $\sim 2$  MHz. The IF bandwidth of the mixer is  $\sim 400$  MHz; the bandwidth of the A/D converter is  $\sim 1$  MHz. The factor that limits the current measurement speed is the kHz range resonant modes of the STM tube scanner.

**Shot noise thermometry.** In the shot noise measurements, the RF excitation arm of the circuit is removed and the homodyne detector is replaced with a spectrum analyser. As d.c. current is driven through the tunnel junction, the shot noise power is collected in the  $LC$  resonant circuit. This noise power is measured by a spectrum analyser. The averaging time for shot noise thermometry is obtained as follows: we take the background temperature as the sum of the ambient temperature, 300 K, and the noise temperature of our amplifier, 70 K:  $T_{bg} \approx 370$  K. The measurement bandwidth is the bandwidth of the resonant tank circuit. In this experiment, the bandwidth is  $B \approx 10$  MHz (HWHM) as shown in Fig. 2a. The averaging time  $\tau$  to achieve  $\Delta T \approx 1$  K temperature resolution in the noise power measurement can be estimated using  $\Delta T/T_{bg} = 1/(B\tau)^{1/2}$  as  $\sim 10$  ms.

**Displacement detection.** In the displacement measurements, we first calibrate the piezoelectric actuator disk motion amplitude using Michelson interferometry:  $dz/dV \approx 0.33\text{ nm V}^{-1}$  at the motion frequency  $f_M \approx 1$  MHz. In the RF-STM-based measurements, the displacement of the Au sample (shaken by the piezo-actuator at frequency  $f_M$ ) is measured by launching an RF (carrier) signal of amplitude  $V_c$  at frequency  $f_{LC}$  towards the tank circuit, and by measuring the reflected signal amplitude  $V_r$ . A time-dependent tunnel junction resistance,  $R_T(t)$ , results in a time-dependent reflection coefficient,  $\Gamma(t)$ , which modulates the amplitude of the reflected signal. As the reflected signal is  $V_r = \Gamma V_c$ , we obtain reflection at  $f_{LC}$  as well as at the sidebands  $f_{LC} + f_M$  and  $f_{LC} - f_M$  (see Fig. 3b).

In membrane displacement measurements, the membranes are driven by a similar piezoelectric disk (Fig. 3a). The piezo-actuator drive is provided by a network analyser, which can perform frequency sweeps near the membrane resonances. The RF-STM signal at  $f_{LC} + f_M$  due to the membrane motion at  $f_M$  is mixed-down with the carrier signal at  $f_{LC}$ . The IF output of the mixer is connected to the input port of the network analyser.



# Fire as the dominant driver of central Canadian boreal forest carbon balance

Ben Bond-Lamberty<sup>1</sup>, Scott D. Peckham<sup>1</sup>, Douglas E. Ahl<sup>1</sup> & Stith T. Gower<sup>1</sup>

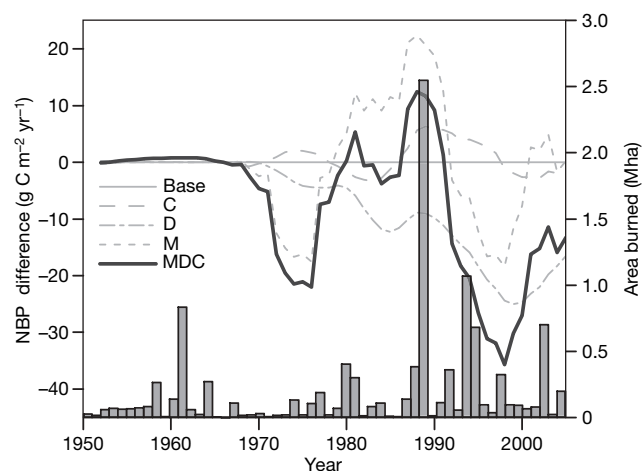
Changes in climate, atmospheric carbon dioxide concentration and fire regimes have been occurring for decades in the global boreal forest<sup>1–3</sup>, with future climate change likely to increase fire frequency<sup>4</sup>—the primary disturbance agent in most boreal forests<sup>3,5</sup>. Previous attempts to assess quantitatively the effect of changing environmental conditions on the net boreal forest carbon balance have not taken into account the competition between different vegetation types on a large scale<sup>6–9</sup>. Here we use a process model with three competing vascular and non-vascular vegetation types to examine the effects of climate, carbon dioxide concentrations and fire disturbance on net biome production, net primary production and vegetation dominance in 100 Mha of Canadian boreal forest. We find that the carbon balance of this region was driven by changes in fire disturbance from 1948 to 2005. Climate changes affected the variability, but not the mean, of the landscape carbon balance, with precipitation exerting a more significant effect than temperature. We show that more frequent and larger fires in the late twentieth century resulted in deciduous trees and mosses increasing production at the expense of coniferous trees. Our model did not however exhibit the increases in total forest net primary production that have been inferred from satellite data<sup>1,10</sup>. We find that poor soil drainage decreased the variability of the landscape carbon balance, which suggests that increased climate and hydrological changes have the potential to affect disproportionately the carbon dynamics of these areas. Overall, we conclude that direct ecophysiological changes resulting from global climate change have not yet been felt in this large boreal region. Variations in the landscape carbon balance and vegetation dominance have so far been driven largely by increases in fire frequency.

We used the Biome-BGC process model<sup>11,12</sup> to simulate the 1,000 km × 1,000 km Boreal Ecosystem-Atmosphere Study (BOREAS) region; at 15–20% of the North American boreal forest, it encompasses much of Manitoba and Saskatchewan as well as eastern Alberta, Canada. Species diversity is low, and the landscape is dominated by black spruce (*Picea mariana* (Mill.) BSP), trembling aspen (*Populus tremuloides* Michx.) and bryophytes. Regional climate changes over the past 60 years have been modest (Supplementary Information). We used a full-factorial design to examine the effects of climate (meteorology, 'M'), CO<sub>2</sub> ('C') and fire disturbance ('D') on net biome production (NBP, net ecosystem carbon balance integrated over time and space<sup>13</sup>), net primary production (NPP), and vegetation dominance. The model included interacting carbon, nitrogen and water cycles<sup>11,14</sup>; a Farquhar photosynthesis model; dynamic and competing vegetation types, including coniferous trees, deciduous broad-leaf trees and mosses<sup>12</sup>; and explicit representation of non-vascular plants<sup>15</sup>, which constitute an important component of boreal biogeochemical cycling.

For the period 1948–2005, mean simulated NBP was  $9.9 \pm 11.8 \text{ g C m}^{-2} \text{ yr}^{-1}$  ( $N = 579,395$  cells, excluding lakes, agricultural

areas, and so on; the error term is interannual variability), indicating the landscape was a small carbon sink until late in the twentieth century. Spatial variability (cell standard deviation) was  $78.9 \text{ g C m}^{-2} \text{ yr}^{-1}$ . The difference between this (the 'MDC' scenario) and what would have happened without any changes in climate, CO<sub>2</sub> or fire (the 'base' scenario), here denoted  $\text{NBP}_{\text{MDC-base}}$ , was  $-423 \text{ g C m}^{-2}$  or  $-8.0 \text{ g C m}^{-2} \text{ yr}^{-1}$  (Fig. 1). Over the final ten years (1996–2005),  $\text{NBP}_{\text{MDC-base}}$  was  $-13.3 \text{ g C m}^{-2} \text{ yr}^{-1}$ . In other words, the simulation indicated that the region lost on average  $423 \text{ g C m}^{-2}$  (or 245 Tg for the entire region, about 2% of soil carbon) over the past 58 years, relative to what would have happened if mid-century conditions had continued unaltered, with the difference increasing in recent years.

The CO<sub>2</sub> effect exerted, by itself, a positive effect on 58-yr mean NBP ( $0.5 \text{ g C m}^{-2} \text{ yr}^{-1}$ ), while the disturbance and meteorology effects were negative ( $-8.5$  and  $-0.3 \text{ g C m}^{-2} \text{ yr}^{-1}$ , respectively). In a sinusoidal model fitted to the NBP output data, the disturbance effect (that is, scenarios D, DC, MD, MDC) changed the wave's offset (mean) by  $-0.3 \text{ g C m}^{-2} \text{ yr}^{-1}$  (Table 1); the meteorology and disturbance effects both increased NBP wave amplitude by  $\sim 0.2 \text{ g C m}^{-2} \text{ yr}^{-1}$ . The CO<sub>2</sub> effect was not significant except interacting with disturbance, where it significantly affected wave period. This interactive effect emphasizes the limitations of extrapolating ecosystem-scale changes from studies on individual forcing effects.



**Figure 1 | Difference in net biome production (NBP) between simulation scenarios, over length of simulation.** A negative value indicates a relative carbon source to the atmosphere. Scenario labels indicate main effects applied to base case (mid-century conditions): climate (meteorology, 'M'), CO<sub>2</sub> ('C'), disturbance ('D') and their combination ('MDC'). Data are smoothed using a 5-yr running mean. Bar graph, on right-hand axis, shows annual area burned in the study area.

<sup>1</sup>Department of Forest and Wildlife Ecology, University of Wisconsin, 1630 Linden Drive, Madison, Wisconsin 53706, USA.

**Table 1 | Significant parameters in sinusoidal model of net biome production**

Parameter	Description	Estimate	s.e.	T	P
$\beta_{1D}$	Disturbance effect on $\beta_0$ (offset) trend	-0.29	0.02	-12.78	<0.0001
$\beta_2$	Wave amplitude	-11.04	1.02	-10.78	<0.0001
$\beta_{3M}$	Meteorology effect on $\beta_2$ trend	0.25	0.04	6.18	<0.0001
$\beta_{3D}$	Disturbance effect on $\beta_2$ trend	0.19	0.04	4.79	<0.0001
$\beta_4$	Wave period	18.88	0.34	55.87	<0.0001
$\beta_{5M}$	Meteorology effect on $\beta_4$ trend	0.04	0.01	4.60	<0.0001
$\beta_{5MDC}$	D×C interaction effect on $\beta_4$ trend	0.31	0.02	14.71	<0.0001
$\phi$	Initial wave phase	6.07	0.55	11.07	<0.0001

See equation (1) in Methods. Estimate (in  $\text{g C m}^{-2}$  for  $\beta_2$ ; in  $\text{g C m}^{-2} \text{yr}^{-1}$  for  $\beta_{1D}$ ,  $\beta_{3M}$ ,  $\beta_{3D}$ ; in yr for  $\beta_4$ ,  $\phi$ ; in  $\text{yr yr}^{-1}$  for  $\beta_{5M}$ ,  $\beta_{5DC}$ ), standard error (s.e.), T statistic and P value are given. D×C refers to two-way interaction of disturbance and  $\text{CO}_2$ . Model residual s.e. was 10.73 on 456 degrees of freedom.

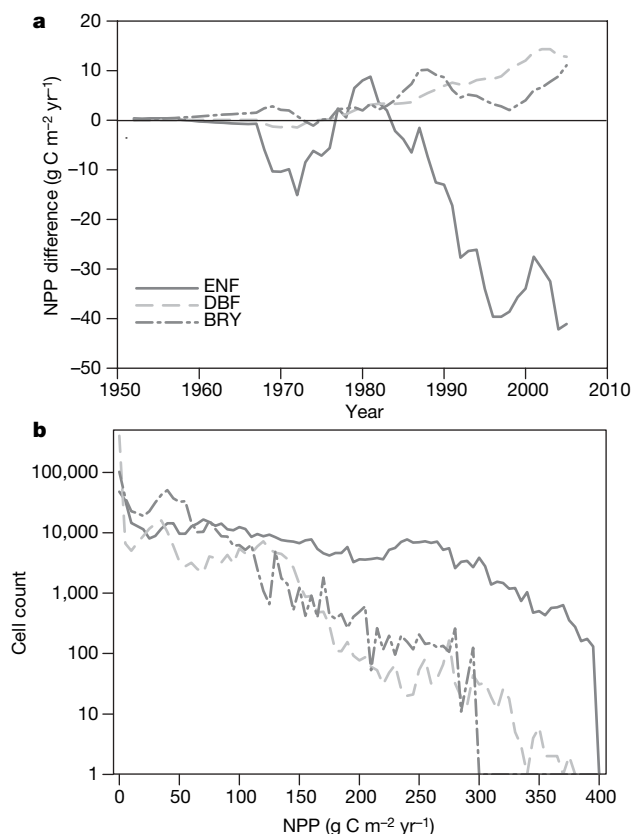
The major and consistent driver of NBP was thus fire disturbance. For the entire grid, mean loss to fire in the base scenario was  $2.7 \pm 0.6 \text{ g C m}^{-2} \text{yr}^{-1}$ , increasing in the real-world MDC scenario to  $6.8 \pm 1.0 \text{ g C m}^{-2} \text{yr}^{-1}$  (spatial variability was respectively 98.7 and  $108.9 \text{ g C m}^{-2} \text{yr}^{-1}$ ). The southwest grid quadrant's NBP was the most variable in time and space ( $23.2$  and  $100.9 \text{ g C m}^{-2} \text{yr}^{-1}$ , respectively) and the northeast the lowest ( $8.9$  and  $46.9 \text{ g C m}^{-2} \text{yr}^{-1}$ , respectively). For the final 30 yr of the simulation, the northwest grid quadrant was a C source ( $-15.1 \text{ g C m}^{-2} \text{yr}^{-1}$ ) while the other three quadrants were small sinks ( $3.9$ – $11.6 \text{ g C m}^{-2} \text{yr}^{-1}$ ). Soil drainage also significantly affected carbon dynamics. NBP was more spatially variable in well drained than in poorly drained areas, and over the final 30 yr of the simulation more variable in time as well (year-to-year standard deviation of 12.2 versus  $11.4 \text{ g C m}^{-2} \text{yr}^{-1}$ , respectively). Thus during this period poorly drained areas damped regional NBP swings (although early in the simulation these areas exhibited

more variability). NBP was positively affected by temperature ( $T_{111} = 4.96$ ,  $P < 0.0001$ ); poorly drained NBP was positively and extremely strongly affected by precipitation, while well drained areas were negatively affected ( $P < 0.0001$  for each).

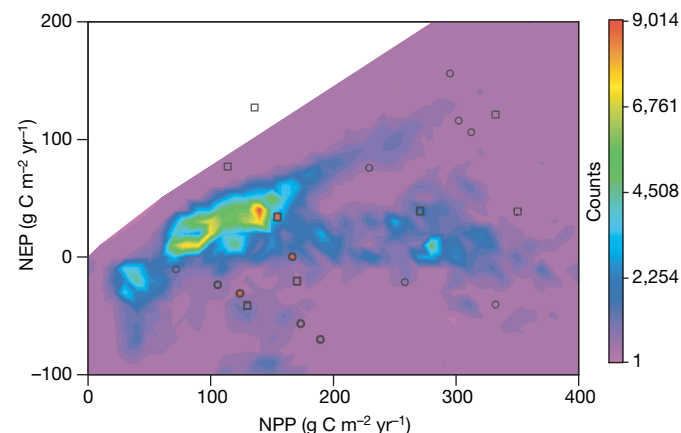
Mean total NPP ( $186.1 \text{ g C m}^{-2} \text{yr}^{-1}$ ) was dominated by evergreen needle-leaf trees ( $141.4 \text{ g C m}^{-2} \text{yr}^{-1}$ ). Conifer NPP declined by  $0.5 \text{ g C m}^{-2} \text{yr}^{-2}$  ( $T_{113} = -7.28$ ,  $P < 0.0001$ ) in the MDC scenario, relative to the base case, and its share of total NPP dropped from 80% at the beginning of the simulation to 73% at its end as fire converted mature coniferous stands to young deciduous ones (Fig. 2a). The distribution of NPP among grid cells was skewed, particularly for deciduous broadleaf trees, which did not occur in 77% of the forest cells; coniferous trees were much more evenly distributed (Fig. 2b). Mean  $\text{NPP}_{\text{MDC}}$  for deciduous broadleaf trees was thus low,  $17.7 \text{ g C m}^{-2} \text{yr}^{-1}$ , and rose by  $0.2 \text{ g C m}^{-2} \text{yr}^{-2}$  between the base and MDC scenarios. Bryophyte NPP averaged  $29.2 \text{ g C m}^{-2} \text{yr}^{-1}$  and increased between scenarios by  $0.1 \text{ g C m}^{-2} \text{yr}^{-2}$ . Mean stem and leaf carbon were respectively 1,785 and  $284 \text{ g m}^{-2}$  in the MDC scenario, while mean deep soil carbon was  $19.8 \text{ kg m}^{-2}$  (12.7 and  $31.6 \text{ kg m}^{-2}$  in the well and poorly drained areas, respectively).

The Biome-BGC model has been extensively validated in North American boreal ecosystems at the stand level (for example, in simulating post-disturbance forest dynamics<sup>15,16</sup>) and in comparison to eddy covariance data<sup>17</sup>. Figure 3 shows how model output compared to results from several chronosequence studies performed at the BOREAS northern and southern study areas in Thompson, Manitoba, and Candle Lake, Saskatchewan, respectively. (We note however that these studies do not constitute a random sample—on the contrary, many were performed in ‘end-member’ stands that constitute a small fraction of the landscape at any given time<sup>18</sup>.)

Comparisons at the biome level, particularly with studies using complementary methods at large temporal and spatial scales, are more appropriate here. The steady-state biomass and soil C values reported above are close to those inferred from inventory and mass-balance modelling<sup>18</sup>. Such inventory approaches have yielded net ecosystem production and NPP estimates of  $-14$  to  $98$  and  $215$  to  $330 \text{ g C m}^{-2} \text{yr}^{-1}$ , respectively, for western boreal Canada, and emphasized the role of natural disturbances<sup>18,19</sup>. Mean carbon lost in each fire event in this study was  $1.5 \pm 1.1 \text{ kg C m}^{-2}$ , comparable to observed data<sup>18,20</sup>, long-term mass balance results<sup>21</sup> and previous modelling efforts<sup>9,22</sup>. We assumed that the boreal landscape was a small C sink in 1948, the beginning of the simulation; inventory models support this assumption<sup>18</sup>, and poorly drained areas in



**Figure 2 | Temporal and spatial distribution of simulated net primary production (NPP).** **a**, Difference in NPP between real-world ‘MDC’ scenario and the mid-century base case, by vegetation type (evergreen needle-leaf trees, ENF; deciduous broadleaf trees, DBF; and bryophytes, BRY) over the simulation. Data are smoothed using a 5-yr running mean. **b**, Histogram of simulated NPP for MDC scenario, all grid cells, year 2005.



**Figure 3 | Comparison of model output with observed data from a variety of studies.** (Studies are listed in Supplementary Information.) Contour plot shows number of grid cells from real-world ‘MDC’ scenario, year 2005, for each combination of net primary production (NPP) and net ecosystem production (NEP). Scatter plot points are published data: circles, data from the BOREAS Northern Study Area; squares, Southern Study Area data. Red symbols are the main BOREAS flux tower sites.

particular had probably not achieved a postglacial steady state<sup>23</sup>. This mean NBP was higher than the long-term sequestration mean implied by soil stocks<sup>21</sup>, but lower than values implied by inverse modelling efforts<sup>24</sup>. Our simulations did not show the strong trend in boreal photosynthetic and total NPP activity inferred from satellite data<sup>1,25,26</sup>: here deciduous broadleaf and bryophyte NPP increased, while evergreen NPP (and total NPP) declined as older stands burned with increasing frequency. Because 5% or more of the landscape can burn in high-fire years, and because many such years have occurred in the last two decades, at regional scales satellites may be recording high post-fire growth, rather than growth directly caused by CO<sub>2</sub> or climate change<sup>27</sup>.

Several sources of uncertainty are worth noting. First, the possible current damping role of poorly drained areas, their large carbon storage, and the control of soil drainage on fire emissions<sup>21</sup> all argue for a better understanding of such systems, but they are poorly modelled by most terrestrial models, including, despite significant improvements<sup>15</sup>, Biome-BGC. We believe this is the largest area of uncertainty with regard to the results presented here. Second, the 58-yr period simulated here is considerably shorter than fire return intervals of the North American boreal forest, and the long-term effects of increased fire may be different from the short-term ones portrayed here. The effects of increased fire will be felt over multiple timescales: short (increased C volatilization), medium (the large areas burned in 1989–95 are only now, 12–18 yr later, turning from C sources to sinks<sup>28</sup>), and long, with shifts in soil carbon stocks, forest age structure, and species dominance.

Fire activity is driven by local and regional climate<sup>2,3</sup>, even if in this study they are treated as separate effects in order to understand quantitatively the contribution of different climate-driven ecosystem processes to forest carbon balance. Fuel loads are a function of long-term climate, while daily weather determines fire activity by drying out the forest fuels, igniting these fuels, and controlling fire spread. Thus climate is indirectly but ultimately responsible for determining the twentieth-century boreal forest carbon dynamics in central Canada. Our results suggest, however, that the ecophysiological changes from global climate change have not yet been felt in this large boreal region; rather, variations in landscape carbon balance and vegetation dominance have so far been largely driven by increases in fire frequency. This implies that fire management may have underappreciated benefits, even if practiced selectively, in affecting decadal boreal carbon balance; currently some degree of fire suppression is attempted on ~50% of large Canadian fires<sup>3</sup>. Our results only apply, however, to systems dominated by stand-replacing fires such as the North American boreal forest. Other regional disturbance agents such as insects and logging are growing in importance<sup>18,29</sup>, and should be considered in future management and modelling efforts.

## METHODS SUMMARY

The processes and logic of Biome-BGC have been described elsewhere<sup>11,12,14</sup>; the version used here was extensively modified with regard to the effects of poor soil drainage<sup>15</sup> and competing vegetation types, including bryophytes<sup>12</sup>. All vegetation types were treated as potentially present in every cell, allowing for species composition changes during simulations. In each fire event, fixed percentages of stem, leaf, coarse root, fine root, litter, woody debris and shallow soil were volatilized, reflecting typical dynamics of a stand-replacing fire. Plant growth was restarted the year following fire. Basic land cover categories were assigned using the MODIS MOD12 data product; soil data were derived from the BOREAS regional soils data. A digital elevation model was obtained from the ORNL DAAC. Meteorological data were downloaded from the Climate Analysis Branch at NOAA for the years 1948–2005. Spatially and temporally explicit data from the Canadian Forest Service's Large Fire Database (LFDB) were used to drive the model; these data include 97% of the area burned in Canada<sup>3</sup>. Plant physiological parameters were standardized across the entire simulation grid and published previously<sup>15</sup>. Simulation cell size was 1 km<sup>2</sup>; each cell was self-initialized until just before its deep soil carbon stabilized (that is, leaving a small ongoing carbon sink). Eight 58-yr simulations, each a combination of the three main effects (meteorology, disturbance and CO<sub>2</sub>) being tested, were then run. Mid-century meteorological and fire data consisted of the 1948–67

time period, cycled as necessary to fill out a 58-yr simulation; changed conditions consisted of the actual 1948–2005 record. Mid-century CO<sub>2</sub> was 310 p.p.m., while in changed conditions it rose from 310 to 380 p.p.m. The carbon balance output by the model was analysed in R 2.4.0 (ref. 30) using a nonlinear sinusoidal model.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 19 December 2006; accepted 6 September 2007.

- Goetz, S. J. *et al.* Satellite-observed photosynthetic trends across boreal North America associated with climate and fire disturbance. *Proc. Natl Acad. Sci. USA* **102**, 13521–13525 (2005).
- Gillett, N. P. *et al.* Detecting the effect of climate change on Canadian forest fires. *Geophys. Res. Lett.* **31**, L18211, doi:10.1029/2004GL020876 (2004).
- Stocks, B. J. *et al.* Large forest fires in Canada, 1959–1997. *J. Geophys. Res.* **108**, 8149, doi:10.1029/2001JD000484 (2003).
- Watson, R. T. *et al.* *The Regional Impacts of Climate Change. An Assessment of Vulnerability* (Cambridge Univ. Press, Cambridge, UK, 1998).
- Zimov, S. A. *et al.* Contribution of disturbance to increasing seasonal amplitude of atmospheric CO<sub>2</sub>. *Science* **284**, 1973–1976 (1999).
- Balshi, M. S. *et al.* Role of historical fire disturbance in the carbon dynamics of the pan-boreal region: a process-based analysis. *J. Geophys. Res.* **112**, G02029, doi:10.1029/2006JG000380 (2007).
- Peng, C. & Apps, M. J. Modelling the response of net primary productivity (NPP) of boreal forest ecosystems to changes in climate and fire disturbance regimes. *Ecol. Modell.* **122**, 175–193 (1999).
- Kang, S., Kimball, J. S. & Running, S. W. Simulating effects of fire disturbance and climate change on boreal forest productivity and evapotranspiration. *Sci. Tot. Environ.* **362**, 85–102 (2006).
- Kasischke, E. S., Christensen, N. L. J. & Stocks, B. J. Fire, global warming, and the carbon balance of boreal forests. *Ecol. Appl.* **5**, 437–451 (1995).
- Nemani, R. R. *et al.* Climate-drive increases in global terrestrial net primary production from 1982 to 1999. *Science* **300**, 1560–1563 (2003).
- Running, S. W. & Coughlan, J. C. A general model of forest ecosystem processes for regional applications. I. Hydrological balance, canopy gas exchange and primary production processes. *Ecol. Modell.* **42**, 125–154 (1988).
- Bond-Lamberty, B. *et al.* Reimplementation of the BIOME-BGC model to simulate successional change. *Tree Physiol.* **25**, 413–424 (2005).
- Chapin, F. S. *et al.* Reconciling carbon-cycle concepts, terminology, and methods. *Ecosystems* **9**, 1041–1050 (2006).
- Running, S. W. & Gower, S. T. FOREST-BGC, A general model of forest ecosystem processes for regional applications. II. Dynamic carbon allocation and nitrogen budgets. *Tree Physiol.* **9**, 147–160 (1991).
- Bond-Lamberty, B., Gower, S. T. & Ahl, D. E. Improved simulation of poorly drained forests using Biome-BGC. *Tree Physiol.* **27**, 703–715 (2007).
- Bond-Lamberty, B. *et al.* Simulation of boreal black spruce chronosequences: comparison to field measurements and model evaluation. *J. Geophys. Res.* **111**, G02014, doi:10.1029/2005JG000123 (2006).
- Amthor, J. S. *et al.* Boreal forest CO<sub>2</sub> exchange and evapotranspiration predicted by nine ecosystem process models: Intermodal comparisons and relationships to field measurements. *J. Geophys. Res.* **D 106**, 33623–33648 (2001).
- Kurz, W. A. & Apps, M. J. A 70-year retrospective analysis of carbon fluxes in the Canadian forest sector. *Ecol. Appl.* **9**, 526–547 (1999).
- Li, Z. *et al.* Temporal changes of forest net primary production and net ecosystem production in west central Canada associated with natural and anthropogenic disturbances. *Can. J. For. Res.* **33**, 2340–2351 (2003).
- Amiro, B. D. *et al.* Direct carbon emissions from Canadian forest fires, 1959–1999. *Can. J. For. Res.* **31**, 512–525 (2001).
- Harden, J. W. *et al.* The role of fire in the boreal carbon budget. *Glob. Change Biol.* **6** (Suppl. 1), 174–184 (2000).
- Kasischke, E. S. & Bruhwiler, L. P. Emissions of carbon dioxide, carbon monoxide, and methane from boreal forest fires in 1998. *J. Geophys. Res.* **108**, 8146, doi:10.1029/2001JD000461 (2003).
- Harden, J. W. *et al.* Dynamics of soil carbon during deglaciation of the Laurentide ice sheet. *Science* **258**, 1921–1924 (1992).
- Deng, F. *et al.* Global monthly CO<sub>2</sub> flux inversion with a focus over North America. *Tellus B* **59**, 179–190 (2007).
- Hicke, J. A. *et al.* Satellite-derived increases in net primary productivity across North America, 1982–1998. *Geophys. Res. Lett.* **29**, 1427, doi:10.1029/2001GL013578 (2002).
- Zhou, L. *et al.* Variations in northern vegetation activity inferred from satellite data of vegetation index during 1981 to 1999. *J. Geophys. Res.* **D 106**, 20069–20083 (2001).
- Hicke, J. A. *et al.* Postfire response of North American boreal forest net primary productivity analyzed with satellite observations. *Glob. Change Biol.* **9**, 1145–1157 (2003).
- Litvak, M. *et al.* Effect of stand age on whole ecosystem CO<sub>2</sub> exchange in the Canadian boreal forest. *J. Geophys. Res.* **D 108**, 8225, doi:10.1029/2001JD000854 (2003).



29. Flannigan, M. D. & Van Wagner, C. E. Climate change and wildfire in Canada. *Can. J. For. Res.* **21**, 66–72 (1991).
30. R Development Core Team. R: A language and environment for statistical computing. (R Foundation for Statistical Computing, Vienna, 2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This research was supported by National Aeronautics and Space Administration and National Science Foundation (Integrated Research Challenges in Environmental Biology) grants to S.T.G. Biome-BGC version 4.1.2

was provided by the Numerical Terradynamic Simulation Group (NTSG) at the University of Montana; NTSG assumes no responsibility for the proper use of Biome-BGC by others. This study would not have been possible without the assistance of D. Luke.

**Author Contributions** S.T.G., B.B.-L. and D.E.A. designed the experiment. S.D.P. and D.E.A. assembled input data, and B.B.-L. and S.D.P. wrote code, ran the model, and analysed output data. B.B.-L. administered the experiment and wrote the manuscript.

**Author Information** Correspondence and requests for materials should be addressed to B.B.-L. ([bpbond@wisc.edu](mailto:bpbond@wisc.edu)).

## METHODS

**Biome-BGC process model.** The processes and logic of Biome-BGC have been described elsewhere<sup>11,12,14</sup>; the version used here was extensively modified with regard to the effects of poor soil drainage<sup>15</sup> and competing vegetation types, including bryophytes<sup>12</sup>. All vegetation types were treated as potentially present in every cell, allowing for species composition changes during simulations. In each fire event, fixed percentages of stem, leaf, coarse root, fine root, litter, woody debris and shallow soil were volatilized, reflecting typical dynamics of a stand-replacing fire. Plant growth was restarted the year following fire.

**Input data sources.** Basic land cover categories were assigned using the MODIS MOD12 data product (<http://www.modis.bu.edu/landcover/userguide/lc.html>); the MODIS MCD43B product was used for albedo: a 16-day composite image from 28 July to 12 August 2003 was calibrated and resampled using bilinear interpolation, and the shortwave band extracted from the 20-band image. MODIS products were imported, subset and masked to include only the simulation grid using nearest neighbour resampling. Soil data were derived from data layers of the BOREAS regional soils data ([http://www.daac.ornl.gov/BOREAS/guides/Region\\_Soils\\_Raster.html](http://www.daac.ornl.gov/BOREAS/guides/Region_Soils_Raster.html)). A digital elevation model was obtained from the ORNL DAAC. All data were reprojected, if necessary, to BOREAS Albers Equal Area Conic.

Estimates of the sand, silt and clay percentages were obtained using the soil texture raster data and the Canadian Soil Information System (CanSIS) website (<http://sis.agr.gc.ca/cansis/>); these were applied to all other soils in the raster image that were similar to known soil textures. Thirty per cent of the BOREAS grid was coded as 'not applicable'. In order to increase the number of pixels with soil texture information, these 'not applicable' locations were filled in by association with land cover. The rooting depth data layer was used to calculate the Biome-BGC effective soil depth parameter, based on four categories (0.2, 0.75, 1.5 and 2.5 m). Values of 'not applicable' were replaced on the basis of soil texture associations.

Our revised version of Biome-BGC has water inflow and outflow rate parameters (inflow  $\alpha_0$ , surface outflow  $\alpha_1$  and subsurface outflow  $\alpha_2$ )<sup>15</sup>. A soil drainage class raster image was used to compute these three BGC parameters based on the following criteria: excessively well drained ( $\alpha_0 = 0.0$ ,  $\alpha_1 = 1.0$ ,  $\alpha_2 = 0.8$ ); well to moderately well drained ( $\alpha_0 = 0.0$ ,  $\alpha_1 = 1.0$ ,  $\alpha_2 = 0.5$ ); poorly drained ( $\alpha_0 = 0.5$ ,  $\alpha_1 = 0.1$ ,  $\alpha_2 = 0.01$ ); very poorly drained ( $\alpha_0 = 1.0$ ,  $\alpha_1 = 0.01$ ,  $\alpha_2 = 0.0$ ).

Meteorological data were downloaded directly from the Climate Analysis Branch at NOAA (<http://www.cdc.noaa.gov/cdc/reanalysis/reanalysis.shtml>) for the years 1948–2005. These global data included maximum daily temperature, minimum daily temperature, specific humidity, shortwave radiation, surface pressure and precipitation rate, and were on the T1 gaussian grid. A subset of

the global grid was chosen to adequately cover the study area, and the data resampled to the simulation grid, with 20 km resolution using weighted distance interpolation with the six nearest latitude/longitude points. Vapour pressure deficit was calculated from surface pressure and specific humidity. In order to make a simple link to cells on the BOREAS grid, a 'met key' grid was created, in which each 1 km<sup>2</sup> cell contained the values of its corresponding meteorology cell. Spatially and temporally explicit data from the Canadian Forest Service's Large Fire Database (LFDB) were used to drive the model; these data include 97% of the area burned in Canada<sup>3</sup>. The LFDB covers 1959–2005; data for 1948–58 were derived on the basis of provincial records of total area burned and the known fire size distributions data of the LFDB. Plant physiological parameters were standardized across the entire simulation grid and published previously<sup>15</sup>.

**Simulation and statistical procedures.** The simulation grid was 10<sup>6</sup> km<sup>2</sup>; simulation cell size was 1 km<sup>2</sup>. Each grid cell was independently self-initialized under historical conditions, with a fire return interval of 175 yr, from an initial empty state until just before its deep soil carbon stabilized (that is, leaving a small ongoing carbon sink). From this point eight 58-yr simulations, each a combination of the three main effects (meteorology, disturbance and CO<sub>2</sub>) being tested, were run. Each effect could be set to either 'mid-century' or 'changed' states. Mid-century meteorological and fire data consisted of the 1948–67 time period, cycled as necessary to fill out a 58-yr simulation; changed conditions consisted of the actual 1948–2005 record. Mid-century CO<sub>2</sub> was 310 p.p.m., while in changed conditions it rose from 310 to 380 p.p.m. All simulations were performed on a distributed computing grid using Apple Computer's Xgrid software to handle job distribution and processing.

Steady-state carbon balance in models such as Biome-BGC tends to oscillate around a mean; this was modelled with a nonlinear model:

$$y = \beta_0 + \beta_1 t + (\beta_2 + \beta_3 t) \sin \left[ \frac{(t - \phi)}{\beta_4 + \beta_5 t} \right] \quad (1)$$

where  $y$  is carbon balance;  $t$  the year, normalized with 1948 as year 0;  $\phi$  the initial wave phase; and  $\beta_0$ ,  $\beta_2$  and  $\beta_4$  the wave offset, amplitude and period, respectively. The three main effects were tested for their separate effects on the linear trends of offset, amplitude and period relative to the detrended base case data. Thus  $\beta_1$ ,  $\beta_3$  and  $\beta_5$  were each the sum of a base value plus main effects and their interactions (that is,  $\beta_i + \beta_{iC} + \beta_{iD} + \beta_{iM} + \beta_{iCD} + \beta_{iCM} + \beta_{iDM} + \beta_{iCDM}$ ; for clarity these are not expanded in equation (1)). This final model had 28 parameters, 464 observations (= 8 scenarios  $\times$  58 years) and was simplified, after initial fitting, by stepwise removal of non-significant terms. Statistical analyses were conducted in R version 2.4.0 (ref. 30), used annual flux or pool as the experimental unit, and a significance level of  $\alpha = 0.05$ .

# Convergent dental adaptations in pseudo-tribosphenic and tribosphenic mammals

Zhe-Xi Luo<sup>1</sup>, Qiang Ji<sup>2</sup> & Chong-Xi Yuan<sup>2</sup>

Tribosphenic molars of basal marsupials and placentals are a major adaptation, with the protocone (pestle) of the upper molar crushing and grinding in the talonid basin (mortar) on the lower molar<sup>1–4</sup>. The extinct pseudo-tribosphenic mammals have a reversed tribosphenic molar in which a pseudo-talonid is anterior to the trigonid, to receive the pseudo-protocone of the upper molar. The pseudo-protocone is analogous to the protocone, but the anteriorly placed pseudo-talonid is opposite to the posterior talonid basin of true tribosphenic mammals<sup>5–7</sup>. Here we describe a mammal of the Middle Jurassic period with highly derived pseudo-tribosphenic molars but predominantly primitive mandibular and skeletal features, and place it in a basal position in mammal phylogeny. Its shoulder girdle and limbs show fossorial features similar to those of mammaliaforms and monotremes, but different compared with those of the earliest-known Laurasian tribosphenic (boreosphenid) mammals. The find reveals a much greater range of dental evolution in Mesozoic mammals than in their extant descendants, and strengthens the hypothesis of homoplasy of ‘tribosphenic-like’ molars among mammals.

Class Mammalia

Clade Yinotheria Chow and Rich, 1982 (emended by ref. 3)

Family Shuotheriidae Chow and Rich, 1982 (ref. 5)

*Pseudotribos robustus* gen. et sp. nov.

**Etymology.** *Pseudo*, false, for superficial resemblance; *tribos*, grinding, for the grinding and crushing function of the pseudo-tribosphenic molar; *robustus*, strong, for the stout limb bones of the new mammal. **Holotype.** Chinese Academy of Geological Sciences (CAGS)040811A and CAGS040811B are the part and counterpart, respectively, of a partial skeleton with impression and carbonized residues of furs (Fig. 1).

**Locality and age.** Daohugou locality (41° 18.979' N, 119° 14.318' E), Ningcheng County, Inner Mongolia Region, China, in the Middle Jurassic Jiulongshan Formation. Mammaliaforms from this formation include the docodont *Castorocauda*<sup>8</sup> and the basal mammal *Volaticotherium*<sup>9</sup> (see also Supplementary Information).

**Diagnosis.** Dental formula:  $I^{2+}-C^1-P^5-M^3/I_4-C_1-P_5-M_3$  (Fig. 2). Upper molars have a triangular pattern of three cusps: a taller paracone, a shorter metacone and an elevated ‘pseudo-protocone’. Lower molars have a triangular cusp pattern of the trigonid with a labial protoconid (tallest), an anterior paraconid and a posterior metaconid (lowest). Anterior to the trigonid is the pseudo-talonid basin with an elevated labial rim (pseudo-hypoconid) that is connected to the pseudo-hypolophid crest. The lingual rim of pseudo-talonid basin of *Pseudotribos* lacks the distinctive cusp ‘pseudo-entoconid’ that is present in the closely related *Shuotherium*. *Pseudotribos* is distinguishable from *Shuotherium* in having a more extensive posterior cingulid extending to the labial side of the trigonid and in having a less developed pseudo-hypoconid (Fig. 3). It is distinguishable from

*Shuotherium shilongi* and *Shuotherium kermacki* in having teeth about 30% smaller in size than comparable teeth, but larger than those of *Shuotherium dongi* by 20%<sup>5–7</sup>. *Pseudotribos* differs from *S. kermacki* in having a more inflated pseudo-protocone. It is distinguishable from other mammaliaforms in the apomorphies of a well developed pseudo-talonid basin positioned anterior to the trigonid on the lower molar in full occlusion with the upper pseudo-protocone (Fig. 2e). It is distinguishable from all therii-morph mammals (from eutriconodontans, multituberculates and spalacotheroids to living therians)<sup>10</sup> in the following plesiomorphies: a large interclavicle twice as long as the sternal manubrium, with two lateral processes and an expanded posterior process. It is distinguishable from multituberculates to living therians (except *Akidolestes*)<sup>11</sup> in retaining mobile lumbar ribs, in lacking a distinctive head–neck offset from the femoral shaft, and in the presence of an enlarged lesser trochanter and a laterally projecting greater trochanter. Although similar to *Fruitafossor* in generalized and robust limb bones, *Pseudotribos* lacks the tubular postcanines and xenarthran-like lumbar bars of the latter<sup>12</sup>. (For a full diagnosis, see Supplementary Information.)

The most important molar feature of the reversed tribosphenic pattern is the pseudo-talonid with a pestle-to-mortar crushing function<sup>5–7</sup>—a convergence to the true tribosphenic molar that is correlated with the great diversification of basal placentals and marsupials<sup>1–4</sup>, and with basal australosphenid mammals<sup>3,13–16</sup> (see Fig. 2). Previous assignment of isolated upper molars to the lower teeth of *Shuotherium*<sup>6,7</sup> is now validated by the *in situ* occlusion of upper and lower teeth of the new *Pseudotribos* (Fig. 2e): pseudo-protocones of left  $M^1$  and  $M^2$  are preserved in occlusion with the anterior pseudo-talonid basins of lower  $M_2$  and  $M_3$ , respectively. The pseudo-protocone of  $M^3$  is smaller and asymmetrical; its prevallum surface shears against the postvallid surface of the trigonid of  $M_3$ ; and the post-cingulid of lower  $M_3$  seems to be too small to have any crushing function (Fig. 2e), as previously predicted<sup>5</sup>. The preserved postvallum of upper  $P^5$  corresponds to the anterior side (prevallid) of the trigonid on lower  $M_1$ . This matches the triangular lower ultimate premolar ( $P_5$ ), as earlier interpreted for *Shuotherium* and some australosphenid mammals<sup>3,13,14</sup>, and aligns all upper and lower molars. The canine is incisiform, and the ante-molar series shows a posterior gradient of increasingly developed precingulids that further expands into the pseudo-talonid basin on  $M_1$ . Conspicuous spaces are present between teeth from the posterior incisors to the lower premolar  $P_4$ . The replacing teeth at the ultimate lower incisor locus and the upper canine locus indicate a typical diphyodont replacement as in most mammaliaforms<sup>3</sup>.

The mandible of *Pseudotribos* is poorly preserved, but reveals several structures previously unknown in shuotheriids (Fig. 2g). The dentary condyle is robust. The mandibular ramus is gracile and shallow in the symphyseal region, but more robust and deeper

<sup>1</sup>Carnegie Museum of Natural History, Pittsburgh, Pennsylvania 15213, USA. <sup>2</sup>Chinese Academy of Geological Sciences, Beijing 100037, China.



(1.6 mm) below  $M_2$  in *Pseudotribos* than in *Shuotherium* (1.3 mm). The mandibular angle in *Pseudotribos* is better developed than in *Shuotherium*, and has a posteriorly facing facet for the ectotympanic (angular) bone as in docodontans<sup>9,17</sup>. A rugose area is tentatively interpreted as the coronoid scar (Fig. 2g). *Pseudotribos* is similar to *Shuotherium* in having a convex ventral border of the mandible under  $M_2$ . The postdentary trough extends anteriorly to the posterior opening of the mandibular canal. The shallow Meckel's sulcus extends anteriorly to below  $M_1$ , and is parallel to the ventral border of the mandible, not intersecting the ventral border as in *Shuotherium*. Judging from the mandibular and molar differences within shuotheriids, there seems to be a wide range of morphological divergence between *Pseudotribos* and *Shuotherium* as well as between known species of *Shuotherium*, suggesting a greater diversity than shown by the currently modest record of this group.

*Pseudotribos* has 22 thoraco-lumbar vertebrae (Fig. 1), including 13 thoracic vertebrae having the ribs connected by means of costal cartilage to the sternbrae. Three additional thoraco-lumbar vertebrae (Fig. 1, tr14–16) have floating ribs without connection to the sternal series, and are tentatively designated thoracic vertebrae. The last six thoraco-lumbar vertebrae (17–22) have more vertical pre-zygapophyses and more robust centra; these are designated lumbar vertebrae. The last three thoracic vertebrae and the anterior two lumbar vertebrae with mobile ribs form a gradation of successively shorter ribs in more posterior vertebrae, as in modern monotremes, the basal mammal *Fruitafossor* and the eutriconodont *Yanoconodon*<sup>12,18</sup>. The costal cartilages of posterior ribs are gracile and individualized, unlike the broad and overlapping (imbricating) costal cartilage plates of extant monotremes<sup>19</sup>. The cervical vertebrae are obscured by the broken elements of the shoulder girdle, but



**Figure 1 | Skeleton of the new Mesozoic mammal *Pseudotribos robustus*.** **a**, Holotype CAGS040811A (main part). **b**, CAGS040811B (counter-part). An outline for identification of main skeletal features is shown to the left of **a** and **b**. **c**, Restoration of *Pseudotribos* as a fossorial mammal with sprawling limbs (grey indicates the bones preserved in the holotype). ap, angular process of mandible; C, canine; cl(in), clavicle (incomplete); cos?, putative coronoid scar (dentary); dc, dentary condyle; ec, ectepicondyle (humerus); ep, epipubis; fe, femur; fi, fibula; fr, frontal; gt, greater trochanter (femur);

hu, humerus;  $I^{1-4}$  and  $I_{1-4}$ , upper and lower incisors 1–4 (or alveoli), respectively; ic, interclavicle; il, ilium; l4–6, lumbar vertebrae 4 to 6; lr1–3, lumbar ribs 1 to 3; lt, lesser trochanter (femur); mb, manubrium (sternbra 1); Mm, upper and lower molars (in partial occlusion); mx, maxilla; oc, occipital bones (outline only); P, upper premolars<sup>5</sup> or their loci indicated by roots; pa, parietal (outline only); ra, radius; s1–3, sacral vertebrae 1–3; sc(in), scapula (incomplete); stb2–7, sternbrae 2–7; t16, thoracic centrum 16; ti, tibia; tr1, thoracic rib 1; tr14–16, thoracic ribs 14–16; ul, ulna.

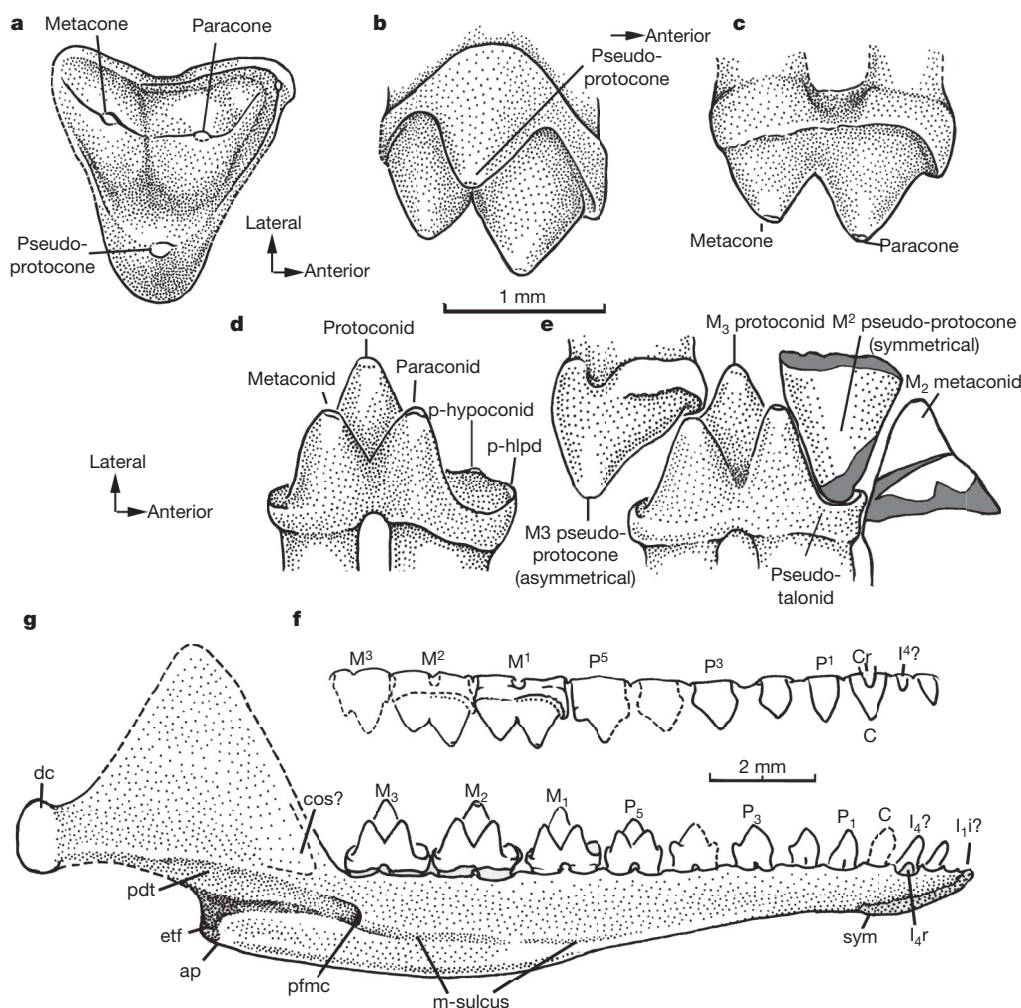
three sacral vertebrae are recognizable by the iliosacral contacting surfaces.

The preserved parts of pelvic and hindlimb bones are robust. The ilium, the femur and the preserved proximal parts of the tibia and fibula are very broad and are thick for their length. An impression overlapping the ilium is interpreted to represent the epipubis, a mammaliaform plesiomorphy<sup>9</sup>. The femoral head is large and without a distinctive neck. The greater trochanter is laterally oriented and larger than the medially directed lesser trochanter below the femoral head. The lateral condyle is larger than the medial condyle. The proximal end of the fibula is broad; the parafibular process is present, but not strongly developed. The tibia has a strong anterior crest, a broad trough for the tibialis anterior muscle, and a proximo-lateral process. These features are plesiomorphic for mammaliaforms and tritylodontids<sup>9,20–23</sup>.

The robust shoulder girdle of *Pseudotribos* shows many fossorial features that are also plesiomorphies for crown Mammalia. This supports a basal placement of shuotheriids in mammaliaform phylogeny. *Pseudotribos* resembles *Sinoconodon* and modern monotremes (Fig. 4) in that the interclavicle is twice the length of the manubrium (cynodont plesiomorphy), and has a constricted waist, a club-foot

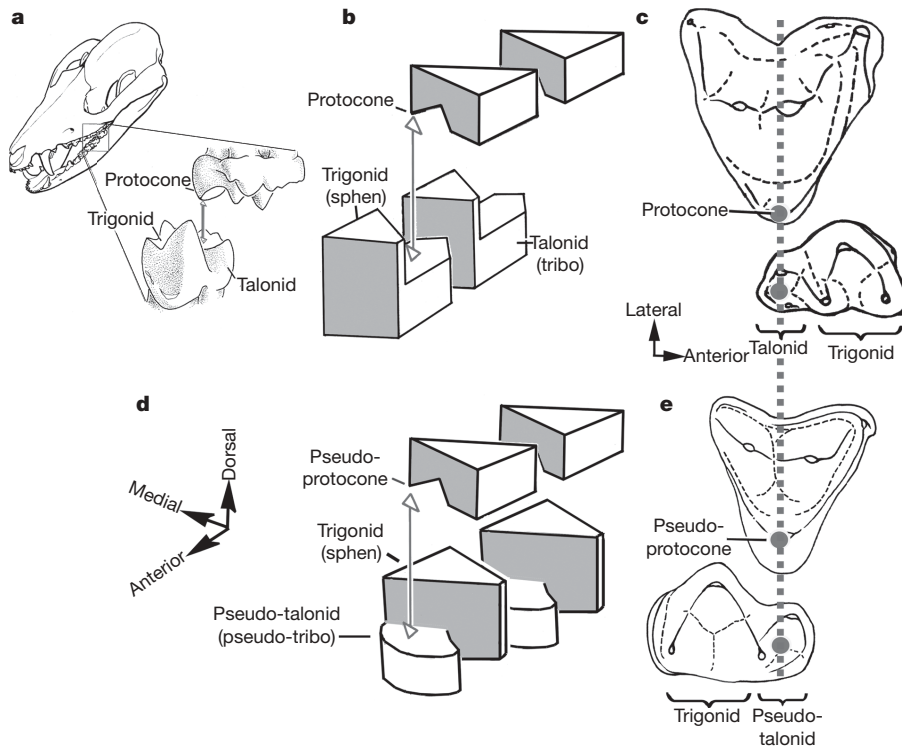
posterior expansion, and elongate articulation for the clavicle (all mammaliaform apomorphies). Tritylodontids have an elongate but transversely narrow interclavicle with a tapering posterior end<sup>22,23</sup> (Fig. 4b)—a basal cynodont condition<sup>24,25</sup> (Fig. 4). In contrast, the posterior end of the interclavicle is expanded and juxtaposed with the expanded anterior side of the ‘T-shaped’ or triangular manubrium (the first rib-bearing element in the sternal structure) in mammaliaforms. The interclavicle and manubrium together form a massive area for the pectoralis muscles. The lateral processes of interclavicle and clavicles form the expanded area of origin for the sternocleidomastoid muscle. The expanded muscular attachments suggest a powerful forelimb with a sprawling posture, as in monotremes with burrowing adaptation<sup>19</sup>, consistent with such fossorial features as the expanded deltopectoral crest, the teres major tubercle, and an expanded distal end of the humerus. The olecranon process is long, and its length is 62% that of the ulnar length anterior to the semilunar notch, similar to *Fruitafossor* and extant fossorial mammals<sup>12</sup>.

In all mammaliaforms for which the pectoral girdle is sufficiently known<sup>12,20,21</sup>, the coracoid does not articulate with the interclavicle. Extant monotremes, however, have a hypertrophied (meta)coracoid



**Figure 2 | Dental and mandibular features of *Pseudotribos*.** **a**, Upper molar  $M^2$ , crown view, composite restoration, the cingular outline conjectural, based on CAGS040811A, CAGS040811B and ref. 7. **b**, Lingual view (with a ventral tilt, based on camera lucida drawing of CAGS040811A). **c**, Labial view (CAGS040811A). **d**, Lower molar  $M_2$ , lingual view, composite restoration from CAGS040811A and CAGS040811B. **e**, Upper molars ( $M^2$ ,  $M^3$ ) in occlusion with lower  $M_2$  and  $M_3$  in lingual view (intact in CAGS040811B). **f**, Composite restoration of the upper teeth, lateral view, only two posterior

incisors known, based on CAGS040811A and CAGS040811B. **g**, Mandible and dentition, medial view, composite restoration of CAGS040811A and CAGS040811B. Cr, replacing (permanent) canine; etf, ectotympanic (angular) facet on the mandibular angle;  $I_1i$ , incisor  $I_1$  alveolus;  $I_4i?$ , possible incisor  $I_4$  alveolus;  $I_4r$  erupting (replacement)  $I_4$ ; m-sulcus, Meckel's groove (dentary); pdt, postdentary trough; pfmc, posterior foramen of mandibular canal; p-hlps, pseudo-hypolophid; p-hypoconid, pseudo-hypoconid; sym, mandibular symphysis.



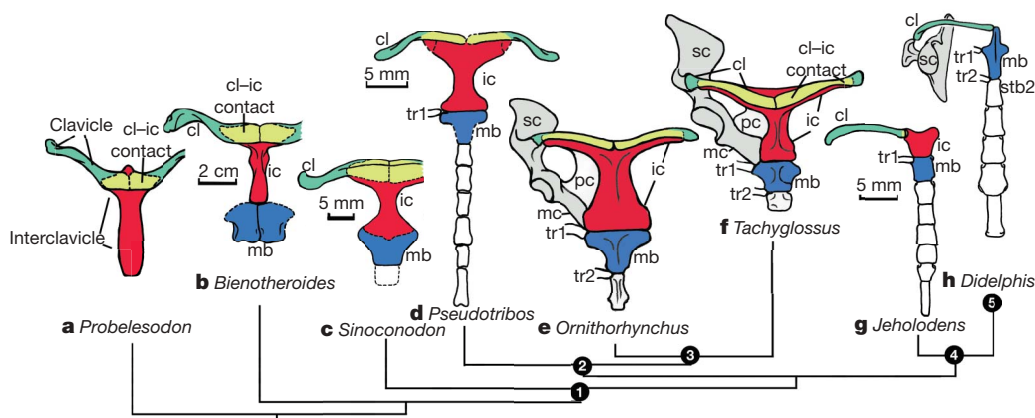
**Figure 3 | Comparison of tribosphenic and pseudo-tribosphenic molars.** **a**, Skull of the marsupial *Didelphis virginiana* with its 'true' tribosphenic molars. **b**, Schematic illustration of the upper-lower occlusion of tribosphenic molars (trigonid anterior to talonid). **c**, Boreosphenidan *Kielantherium* as example of the ancestral tribosphenic pattern<sup>4</sup>.

**d**, Schematic illustration of the upper-lower occlusion of pseudo-tribosphenic molars (trigonid posterior to pseudo-talonid). **e**, Shuotheriids as an example of pseudo-tribosphenic molars (based on *Pseudotribos*). Dashed line shows the protocone-talonid occlusion and the pseudoprotocone-pseudotalonid occlusion.

extending to articulate with the expanded posterior end of the interclavicle (Fig. 4e, f)<sup>19,26</sup>. This provides extra support for the powerful shoulder and forelimb for burrowing in modern monotremes. The posterior club-foot of the interclavicle and the expanded anterior end of the manubrium are also present in *Pseudotribos*. This primitive morphotype of girdle and sternal structure, as seen in *Sinoconodon*,

*Pseudotribos* and monotremes, differs from the pivotal claviculo-interclavicle joint and mobile shoulder girdle of theriomorphs (eutriiconodontans and crown therians) (Fig. 4)<sup>11,18,26–28</sup>.

The reversed tribosphenic molar structure is significant by its opposite arrangement of the two functional parts (the primitive trigonid versus the derived talonid) of the tribosphenic molar of basal



**Figure 4 | Evolution of interclavicle and anterior sternal structure in the cynodont-mammal transition.** (Placement of *Pseudotribos* in mammal phylogeny is shown in Supplementary Information.) **a**, Cynodont *Probelesodon*<sup>25</sup>. **b**, Mammaliaform *Bienotheroides*<sup>22</sup>. **c**, Mammaliaform *Sinoconodon* (restoration from several specimens). **d**, *Pseudotribos* (composite from CAGS040811A and CAGS040811B). **e**, Monotreme *Ornithorhynchus*. **f**, *Tachyglossus*. **g**, Eutriiconodont *Jeholodens*<sup>27</sup>. **h**, Marsupial *Didelphis*. Apomorphies of (1) mammaliaforms: widening anterior end of interclavicle; posterior club-foot of interclavicle to the same width of manubrium for extensive attachment of pectoralis muscles. (2) Yinotheria (*sensu*, ref. 3): 'T-shaped' manubrium. (3) Monotremata: gracile and elongate lateral process of interclavicle overlapping two-thirds the

length of clavicle; enlargement of (meta)coracoid to articulate directly with the interclavicle; presence of procoracoid for strengthening the girdle<sup>19,26,30</sup>. (4) Theriomorpha<sup>10</sup>: shortening of interclavicle to equal manubrium; reduction of the lateral process of interclavicle; mobile claviculo-interclavicle articulation<sup>18,27</sup>. (5) Crown Theria: incorporation of embryonic interclavicle into the manubrium<sup>30</sup>, the first sternal segment in articulation of clavicle and thoracic rib 1. cl, clavicle (green); cl-ic contact, articulation or overlap of clavicle and interclavicle (yellow); ic, interclavicle (red); mb, manubrium (sternebra 1) (blue); mc, metacoracoid; pc, procoracoid<sup>19</sup>; sc, scapula (grey); stb2, second sternebra; tr1 and tr2, thoracic rib costal cartilage 1 and 2, respectively.



marsupials and placentals (Fig. 3) and australosphenid mammals<sup>15,16</sup>. Not present in any modern mammals, this evolutionary phenomenon suggests a much greater range of dental diversity in mammals from the Mesozoic era than in modern mammals.

Despite their importance to mammal dental evolution, shuotheriids were represented previously by only 11 teeth and an incomplete mandible. Because of the conflicting characters of the derived pseudo-tribosphenic molars and primitive mandible represented by its limited fossils, hypotheses about phylogenetic affinities of *Shuotherium* have differed widely. Previously, *Shuotherium* was considered to be intermediate between the primitive mammaliaform *Kuehneotherium* with molar cusps in an obtuse triangle, and the symmetrodonts with molar cusps in an acute triangle<sup>5</sup>, to be related to docodont mammaliaforms<sup>28</sup>, to be between the acute-triangular symmetrodonts and the derived pre-tribosphenic mammal *Peramus*<sup>7</sup>, or to be a basal mammal and sister-group to australosphenidans<sup>3,13–15</sup>. Our analysis on an improved data set from the new *Pseudotribos* excludes an affinity of shuotheriids to symmetrodonts, or to the clade of *Peramus* and more derived lineages. Shuotheriids are more closely related to the australosphenidan clade than other Mesozoic mammal groups<sup>3,13–15</sup>.

Before the discovery of the geometrically reversed design of pseudo-tribosphenic molars, it was widely assumed for more than a century that the upper molar protocone, the lower molar talonid and their occlusal correspondence evolved together in a single origin in early mammalian history. The discovery of *Shuotherium* changed this assumption. It is now generally accepted that a protocone-like structure of the upper molar can occlude either a talonid in the posterior part of lower molar or a pseudo-talonid in the anterior part of the lower molar in different clades; the functionally adaptive upper molar protocone evolved at least twice in mammalian history<sup>4–7</sup>. Recently it was also proposed that the talonid basin in the lower tribosphenic molar has evolved more than once in early mammals<sup>3,4,13–16</sup>. This observation received support from newly discovered australosphenidan mammals<sup>15,16</sup> that lack the typical wear facets of the boreosphenid mammals (as defined by studies for the Laurasian tribosphenic molars)<sup>1,2</sup>. Their detailed occlusal relationship to the structure of upper molars may differ between the Laurasian and the Gondwanan Mesozoic mammals with tribosphenic teeth<sup>15,16</sup>. The new fossil of *Pseudotribos* adds to the inference that both the upper molar protocone and the lower molar talonid could have undergone convergent evolution, and that there is more than one pathway to combine slicing and crushing functions in a single jaw motion for more effective faunivory and omnivory in early mammalian history<sup>29</sup>. The wider morphological range of tooth form and function in Mesozoic mammals suggests that their dental evolution is much more labile than can be inferred from Cenozoic mammals.

Received 3 April; accepted 3 September 2007.

1. Crompton, A. W. in *Early Mammals* (eds Kermack, D. M. & Kermack, K. A.) 65–87 (Zool. J. Linn. Soc., Suppl. 1, London, 1971).
2. McKenna, M. C. in *Phylogeny of the Primates* (eds Luckett, W. P. & Szalay, F. S.) 21–46 (Plenum, New York, 1975).
3. Kielan-Jaworowska, Z., Cifelli, R. L. & Luo, Z.-X. *Mammals from the Age of Dinosaurs — Origins, Evolution, and Structure* (Columbia Univ. Press, New York, 2004).
4. Lopatin, A. V. & Averianov, A. O. An aegialodontid upper molar and the evolution of mammal dentition. *Science* **313**, 1092 (2006).
5. Chow, M. & Rich, T. H. *Shuotherium dongi*, n. gen. and sp., a therian with pseudo-tribosphenic molars from the Jurassic of Sichuan, China. *Austr. Mam.* **5**, 127–142 (1982).
6. Sigogneau-Russell, D. Discovery of a Late Jurassic Chinese mammal in the upper Bathonian of England. *C.R. Acad. Sci. II* **327**, 571–576 (1998).

7. Wang, Y.-Q., Clemens, W. A., Hu, Y.-M. & Li, C.-K. A probable pseudo-tribosphenic upper molar from the Late Jurassic of China and the early radiation of the Holotheria. *J. Vert. Paleontol.* **18**, 777–787 (1998).
8. Ji, Q. et al. A swimming mammaliaform from the Middle Jurassic and ecomorphological diversification of early mammals. *Science* **311**, 1123–1127 (2006).
9. Meng, J. Hu, Y.-M., Wang, Y.-Q., Wang, X.-L. & Li, C.-K. A Mesozoic gliding mammal from northeastern China. *Nature* **444**, 889–893 (2006).
10. Rowe, T. B. Definition, diagnosis, and origin of Mammalia. *J. Vert. Paleontol.* **8**, 241–264 (1988).
11. Li, G. & Luo, Z.-X. A Cretaceous symmetrodont therian with some monotreme-like postcranial features. *Nature* **439**, 195–200 (2006).
12. Luo, Z.-X. & Wible, J. R. A new Late Jurassic digging mammal and early mammalian diversification. *Science* **308**, 103–107 (2005).
13. Luo, Z.-X. et al. Dual origin of tribosphenic mammals. *Nature* **409**, 53–57 (2001).
14. Luo, Z.-X. et al. In quest for a phylogeny of Mesozoic mammals. *Acta Palaeont. Polonica* **47**, 1–78 (2002).
15. Martin, T. & Rauhut, O. W. M. Mandible and dentition of *Asfaltomylos patagonicus* (Australosphenida, Mammalia) and the evolution of tribosphenic teeth. *J. Vert. Paleontol.* **25**, 414–425 (2005).
16. Rougier, G. W. et al. New Jurassic mammals from Patagonia, Argentina: a reappraisal of australosphenidan morphology and interrelationship. *Am. Mus. Novit.* **3566**, 1–54 (2007).
17. Lillegraven, J. A. & Krusat, G. Cranio-mandibular anatomy of *Haldanodon expectatus* (Docodonta; Mammalia) from the Late Jurassic of Portugal and its implications to the evolution of mammalian characters. *Contrib. Geol. Univ. Wyom.* **28**, 39–138 (1991).
18. Luo, Z.-X. et al. A new eutriconodont mammal and evolutionary development in early mammals. *Nature* **446**, 288–293 (2007).
19. Augee, M. L. et al. *Echidna – Extraordinary Egg-laying Mammal* (CSIRO Publishing, Collingwood, 2006).
20. Martin, T. Postcranial anatomy of *Haldanodon expectatus* (Mammalia, Docodonta) from the Late Jurassic (Kimmeridgian) of Portugal and its bearing for mammalian evolution. *Zool. J. Linn. Soc.* **145**, 219–248 (2005).
21. Jenkins, F. A. Jr & Parrington, F. R. The postcranial skeletons of the Triassic mammals *Eozostrodon*, *Megazostrodon* and *Erythrotherium*. *Phil. Trans. R. Soc. Lond. B* **273**, 387–431 (1976).
22. Sun, A.-L. & Li, Y.-H. The postcranial skeleton of Jurassic tritylodonts from Sichuan Province. *Vert. Palaeontol.* **23**, 135–151 (1985).
23. Sues, H.-D. & Jenkins, F. A. Jr. in *Amniote Paleobiology: Perspectives on the Evolution of Mammals, Birds, and Reptiles* (eds Carrano, M. T. et al.) 114–152 (Univ. Chicago Press, Chicago, 2006).
24. Jenkins, F. A. Jr. The postcranial skeleton of African cynodonts. *Peabody Mus. Nat. Hist. Bull.* **36**, 1–216 (1971).
25. Romer, A. S. & Lewis, A. D. The Chañares (Argentina) Triassic reptile fauna. XIX. Post cranial materials of the cynodonts *Probelesodon* and *Probainognathus*. *Breviora* **407**, 1–26 (1973).
26. Sereno, P. C. in *Amniote Paleobiology: Perspectives on the Evolution of Mammals, Birds, and Reptiles* (eds Carrano, M. T. et al.) 315–366 (Univ. Chicago Press, Chicago, 2006).
27. Ji, Q. et al. A Chinese triconodont mammal and mosaic evolution of the mammalian skeleton. *Nature* **398**, 326–330 (1999).
28. Kermack, K. A. et al. A new docodont from the Forest Marble. *Zool. J. Linn. Soc.* **89**, 1–39 (1987).
29. Evans, A. R. & Sanson, G. D. The tooth of perfection: functional and spatial constraints on mammalian tooth shape. *Biol. J. Linn. Soc.* **78**, 173–191 (2003).
30. Klima, M. Early development of the shoulder girdle and sternum in marsupials (Mammalia: Metatheria). *Adv. Anat. Embr. Cell Biol.* **109**, 1–91 (1987).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank A. R. Tabrum for fossil preparation, H.-L. You for his assistance, M. R. Dawson for improving the manuscript, M. A. Klingler for help with graphics, and J. R. Wible, G.-H. Cui and K.-Q. Gao for access to comparative materials. This work was supported by the 973 Project by Ministry of Science and Technology, Chinese Academy of Geological Sciences (Beijing) (Q.J.), the National Science Foundation (USA), the National Natural Science Foundation (China) and the National Geographic Society (Z.-X.L.).

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to Z.-X.L. ([luoz@carnegiemnh.org](mailto:luoz@carnegiemnh.org)).

## LETTERS

# Thresholds and the resilience of Caribbean coral reefs

Peter J. Mumby<sup>1</sup>, Alan Hastings<sup>2</sup> & Helen J. Edwards<sup>1</sup>

The deteriorating health of the world's coral reefs threatens global biodiversity, ecosystem function, and the livelihoods of millions of people living in tropical coastal regions<sup>1</sup>. Reefs in the Caribbean are among the most heavily affected<sup>2,3</sup>, having experienced mass disease-induced mortality of the herbivorous urchin *Diadema antillarum*<sup>4</sup> in 1983 and two framework-building species of coral<sup>5</sup>. Declining reef health is characterized by increases in macroalgae. A critical question is whether the observed macroalgal bloom on Caribbean reefs is easily reversible. To answer this question, we must resolve whether algal-dominated reefs are an alternative stable state of the ecosystem or simply the readily reversible result of a phase change along a gradient of some environmental or ecological parameter<sup>6</sup>. Here, using a fully parameterized simulation model in combination with a simple analytical model, we show that Caribbean reefs became susceptible to alternative stable states once the urchin mortality event of 1983 confined the majority of grazing to parrotfishes. We reveal dramatic hysteresis in a natural system<sup>7,8</sup> and define critical thresholds of grazing and coral cover beyond which resilience is lost. Most grazing thresholds lie near the upper level observed for parrotfishes in nature, suggesting that reefs are highly sensitive to parrotfish exploitation. Ecosystem thresholds can be combined with stochastic models of disturbance to identify targets for the restoration of ecosystem processes. We illustrate this principle by estimating the relationship between current reef state (coral cover and grazing) and the probability that the reef will withstand moderate hurricane intensity for two decades without becoming entrained in a shift towards a stable macroalgal-dominated state. Such targets may help reef managers face the challenge of addressing global disturbance at local scales.

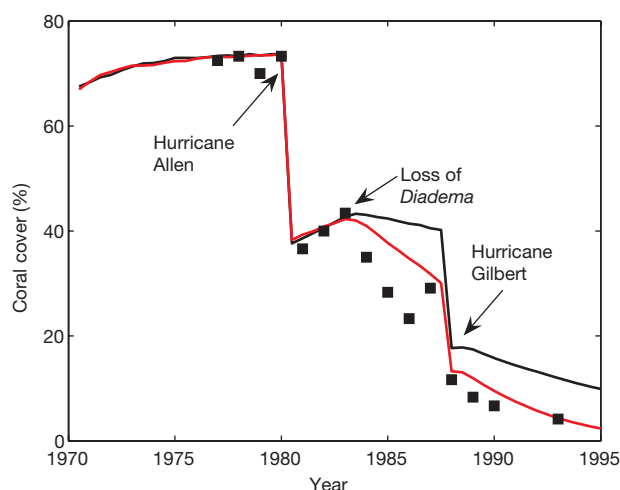
Several studies have documented phase changes from coral- to algal-dominated states on Caribbean reefs<sup>9–11</sup> but none were designed to distinguish simple quantitative changes from the dramatic qualitative changes associated with multiple stable states and hysteresis<sup>12</sup>. Experimental evidence for multiple stable states would need to identify the emergence of multiple stable equilibria from a single parameter value. Given that ethical and logistical issues constrain an experimental approach to this problem<sup>12</sup>, we discover multiple stable equilibria by using a mechanistic model of the ecosystem<sup>13</sup>.

We modelled structurally complex forereef habitats of the Caribbean using a simulation that had the advantage of exploring reef dynamics with a minimum of simplifying assumptions (full details and parameters are in the Supplementary Information). All model parameters were derived from empirical studies in the Leeward Islands<sup>14</sup>, southern Caribbean<sup>15,16</sup> and Central America<sup>17</sup>. Brooding and spawning forms of coral recruit at a size of 1 cm<sup>2</sup> and experience both chronic and acute mortality. Macroalgae, based on the dominant genera *Lobophora* and *Dictyota*, proliferate if dead coral is not sufficiently grazed and have a limited capacity either to arrest coral growth<sup>18</sup> or to overgrow living coral when in direct proximity<sup>19</sup>. An unexploited community of parrotfishes can maintain

approximately 40% of the reef in a permanently grazed state but overfishing reduces this capacity to about 5% (refs 13, 20). Modest urchin populations are more effective grazers than parrotfishes<sup>16</sup>. Comparing model predictions to an exceptionally long time series of independent field data from Jamaica<sup>10</sup>, we find that the model emulates coral dynamics faithfully even when the rate of algal–coral overgrowth is varied within published levels (Fig. 1).

The model suggests that Caribbean coral reefs did not exhibit an algal-dominated stable state when the urchin *Diadema* was present in the system (Fig. 2). At grazing levels exceeding 0.42, meaning that parrotfishes and urchins graze at least 42% of the reef every six months, all reefs, regardless of their state (coral cover), show an upward trajectory towards an equilibrium of high coral cover. However, two stable states emerged after the mass mortality of urchins in 1983 when grazing became dominated by parrotfishes (grazing intensity 0.05–0.4). The open squares in Fig. 2 represent unstable equilibria that join upper and lower stable equilibria. Reefs above and to the right of an unstable equilibrium follow a trajectory towards a stable equilibrium at high coral cover whereas those below and left of the line decline to a stable, macroalgal-dominated state with low coral cover (macroalgal cover not shown).

Multiple equilibria occur because of ecological feedbacks. For example, a decline in coral cover liberates new space for algal colonization. Once maximum levels of grazing have been reached, further increases in grazable area reduce the mean intensity of grazing and increase the probability that a patch of macroalgae will establish,



**Figure 1 | Comparison between model predictions and Hughes' empirical data for the trajectory of structurally complex forereefs in Jamaica at a depth of 10 m.** Predictions are denoted by lines and empirical data are denoted by black squares. The model that was run with a median algal–coral overgrowth rate of 8 cm<sup>2</sup> yr<sup>−1</sup> is shown with a black line and the upper rate of 14 cm<sup>2</sup> yr<sup>−1</sup> is shown with a red line.

<sup>1</sup>Marine Spatial Ecology Lab, School of BioSciences, University of Exeter, Prince of Wales Road, Exeter EX4 4PS, UK. <sup>2</sup>Environmental Science and Policy, University of California, Davis, California 95616, USA.

ungrazed, from the algal turf. The resulting rise in macroalgal cover reduces the availability of coral settlement space and increases the frequency and intensity of coral–algal interactions, thereby reducing coral recruitment, reducing the growth rate of corals and causing limited mortality. The resulting increase in coral mortality further reduces the intensity of grazing, thereby reinforcing the increase in macroalgae. Importantly, hysteresis behaviour appears to be a fundamental property of the ecosystem. To confirm this and determine the features that are essential for the hysteresis effect, we simplified the ecosystem into a three-state analytical model with corals, macroalgae and short algal turfs.

Let  $C$ ,  $T$  and  $M$  represent (as a fraction of the seabed available) the cover of corals, algal turfs and macroalgae respectively. Thus we assume that the sum  $T + M + C$  is constant over time (here assumed to equal one), so only two equations are required to describe the dynamics of the reef (and we can express  $T$  as  $1 - M - C$ ):

$$\frac{dM}{dt} = aMC - \frac{gM}{M+T} + \gamma MT \quad (1)$$

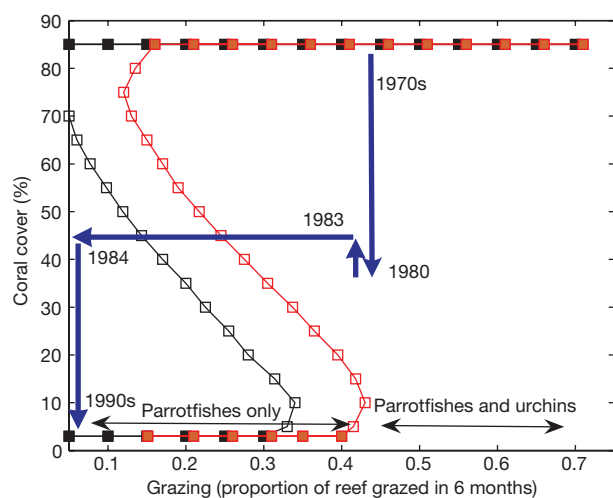
$$\frac{dC}{dt} = rTC - dC - aMC \quad (2)$$

Grazers are assumed not to discriminate between algal types. Algal turfs arise when macroalgae are grazed ( $gM/(M+T)$ ) and as a result of natural coral mortality ( $-dC$ ). Corals recruit to and overgrow algal turfs at a combined rate  $r$ , constrained by the existing cover of turfs. Corals can be overgrown by macroalgae ( $-aMC$ ) but macroalgae usually colonize dead coral by spreading vegetatively over algal turfs ( $\gamma MT$ ).

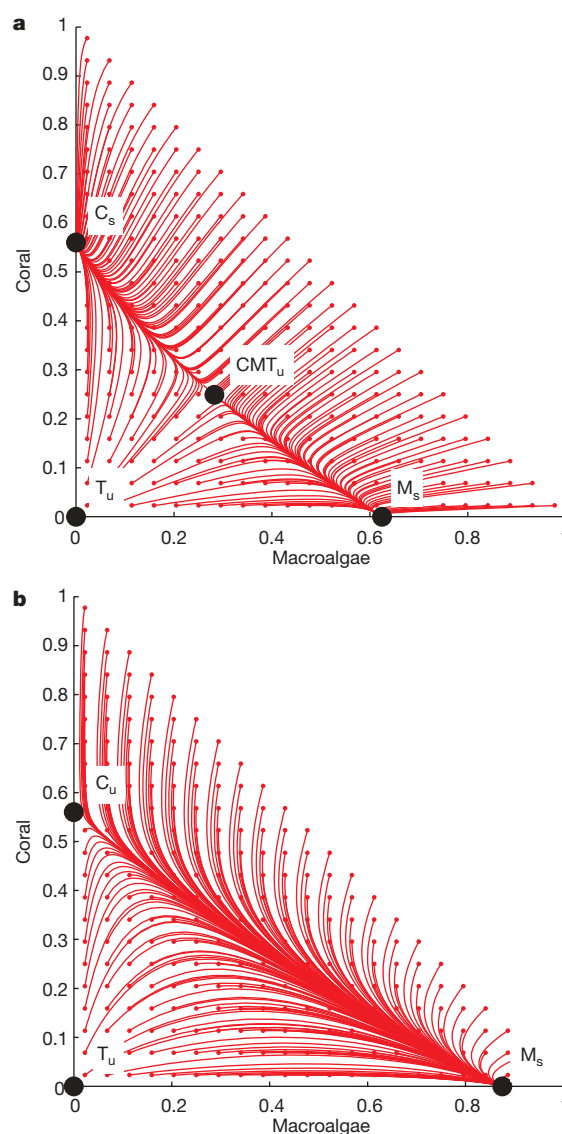
Our analysis of this system has been primarily numerical. The phase plane<sup>21</sup> in Fig. 3 shows all possible dynamical behaviour of the system. Each trajectory describes how populations change over time for a given initial state. All trajectories are attracted to a stable equilibrium, at which point the covers of coral and macroalgae remain constant. The dynamical behaviour of the system differs dramatically between high and low levels of grazing (compare Fig. 3a and Fig. 3b). Two stable states, with either coral ( $C_s$ ) or macroalgal-domination ( $M_s$ ) arise at high grazing, as predicted by the simulation model (Fig. 3a). In addition, there are two non-negative unstable equilibria including a saddle point with corals, macroalgae and turf (CMT<sub>u</sub>), equivalent to the unstable equilibrium represented by the open squares in Fig. 2, and a system dominated by turf ( $T_u$ ), which is never reached when reefs start with positive values of macroalgae or

corals. In contrast, only a single stable equilibrium exists ( $M_s$ ) when grazing levels are low (Fig. 3b) and  $C_s$  becomes an unstable saddle ( $C_u$ ). The analytical model thus replicates the results of the simulation model in that the system may be driven to a high or low coral state and suggests that the essential interactions are those included in the analytical model, thereby confirming our explanation of the dynamics of the simulation model and the natural system.

The most appealing aspect of the hysteresis plot (Fig. 2) is that it helps define reef management objectives by explicitly integrating the impacts of disturbance with the effects of conservation action. The locations of thresholds and bifurcation points reflect the underlying ecosystem dynamics and are influenced by processes such as primary production and chronic coral mortality rates. Acute disturbance phenomena such as coral bleaching<sup>22</sup> cause sudden coral mortality and shift the state of a reef down the  $y$  axis, whereas periods of recovery allow the reef to move back up the  $y$  axis. Changes in grazing on the  $x$  axis represent the effects of fishing herbivores (shifts from right to left) and active herbivore management, such as marine



**Figure 2 | Stable and unstable equilibria for Caribbean coral reefs at two levels of algal–coral overgrowth.** Stable and unstable equilibria are denoted by solid and open squares, respectively. Black denotes  $8 \text{ cm}^2 \text{ yr}^{-1}$ , red denotes  $14 \text{ cm}^2 \text{ yr}^{-1}$ . Blue lines, marked with appropriate dates, represent model predictions of the trajectory of reefs in Jamaica (see also Fig. 1).



**Figure 3 | Phase plane showing equilibrium covers of macroalgae and corals and trajectories over time.** a, b, Equilibrium covers are denoted by black circles; trajectories are shown in red for grazing intensities of 0.3 (a) and 0.1 (b). Trajectories may be thought of as arrows beginning at different initial covers (red circles) and tending towards the stable (subscript s) rather than unstable (subscript u) equilibrium.



reserve implementation or the banning of fish traps (shifts from left to right).

The interpretation of such plots is easily illustrated for the decline in health of some Jamaican reefs between 1981 and 1993 (Fig. 2). By 1979, forereefs had not experienced a severe hurricane for 36 years and coral cover was high at  $\sim 75\%$  (ref. 10). In 1980, a combination of coral disease and hurricane Allen reduced coral cover to around 38%, but because urchins were present, the reef began to recover. When the urchins died out in 1983, grazing levels were decimated, in part because long-term overfishing had removed larger parrotfishes. With a coral cover of approximately 44% and a grazing intensity of only 0.05–0.113, the reef began a negative trajectory towards algal domination that was exacerbated by further acute disturbance. By 1993, coral cover had fallen to less than 5%. A key feature of this graph is that reversing reef decline becomes ever more difficult as the cover of coral declines; as coral cover drops, the level of grazing needed to place the reef on the reverse trajectory (to the right of the unstable equilibrium) increases. Continuing the example of a simulated Jamaican reef, conservation action in the mid-1990s would require grazing levels to be elevated at least fourfold to the maximum observed levels for fishes in the Caribbean. In contrast, if action were taken a decade earlier when coral cover was still around 30%, target grazing levels would be more easily achievable, requiring only a two- to threefold increase.

Hysteresis plots from a field-tested simulation model may improve ecosystem management. Rather than suspecting that bistability occurs at some abstract level of grazing, which might be predicted by an analytical model, it now appears that fishing effects on parrotfish grazing may profoundly influence coral dynamics. Therefore, active management of parrotfishes is both highly desirable and a feasible conservation goal<sup>20</sup>. Recent reports of *Diadema* recovery<sup>23</sup> are promising given the high resilience conferred by the species, but the impact of urchins is only beneficial if their predators are not over-exploited, which would lead to undesirable urchin plagues<sup>24</sup>. Moreover, the hysteresis plot can be combined with stochastic model simulations to create a novel and explicit metric of resilience and set targets for restoration (for example, target levels of grazing). For a given level of grazing, a reef will be resilient if the net impact of disturbance and recovery does not force its state below the unstable equilibrium. This in turn depends on the local disturbance regime<sup>25</sup>. Using stochastic simulation, we can determine the probability that a

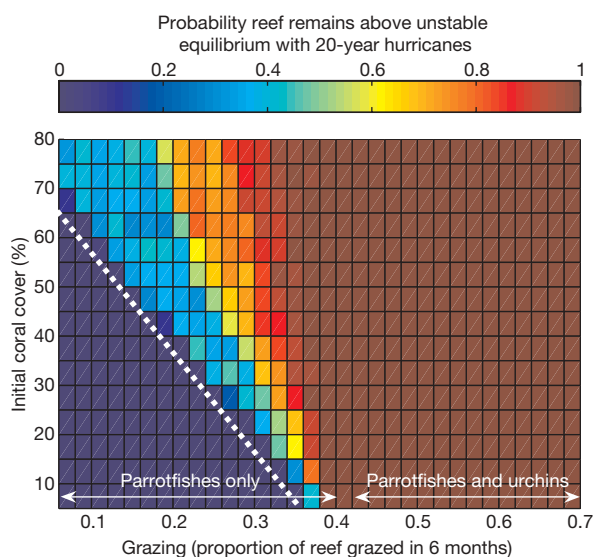
reef will remain above the unstable equilibrium and an example is given for reefs in Central America that experience acute hurricanes with a 20-year periodicity (Fig. 4). In principle, this metric of resilience could be calculated for any physical environment, provided that a parameterization for disturbance events was available. For a given biophysical environment, our approach should help to identify the target levels of coral cover and grazing needed to reverse ecosystem decline and illuminate the efficacy of local conservation policy options against the threat of global climate change on coral reefs.

## METHODS SUMMARY

The simulation model followed the dynamics of individual coral colonies in discrete six-month steps. Although the reef has a toroidal lattice of 2,500 cells (each of size  $0.25\text{ m}^2$ ), the lattice merely helps to define probabilistic rules of coral recruitment and vegetative algal growth. Individual cells comprise multiple coral colonies and algal patches, so interactions occur at colony scales. Unstable equilibria were found by disabling acute disturbance and running simulations with a 25-year period. Initial coral cover was set at 5% intervals with an even mix of brooding and spawning taxa. Grazing was manipulated at intervals of 0.005 and unstable equilibria were defined as combinations of grazing and coral cover for which total cover remained within 2% of its initial level. A resilience plot was generated by conducting 100 simulations of a 25-year run and recording the probability that a reef drops below the mean unstable equilibrium. Hurricanes occurred stochastically but with a long-term 20-year periodicity. Sensitivity analyses reveal that perturbations to parameters and a full stochastic version of the model have little influence on the location of grazing thresholds. The model is most sensitive to the growth rate of brooding corals and changes in the rate of live coral overgrowth by macroalgae, for which extreme values are plotted in Fig. 2. However, variability in either parameter does not alter our conclusions because grazing thresholds are still found to coincide with the upper level of fish grazing.

Equilibria and their stability were determined for the analytic model by standard means<sup>21</sup>. Trajectories were plotted using numerical integration with 4th- and 5th-order Runge–Kutta formulae<sup>26</sup>. We use as simple a model as possible and ignore some time delays that must be present in natural systems (for example, we assume that algal turfs arise instantaneously after macroalgae are grazed). Such assumptions will not affect the equilibrium structure, and should have only limited effect on stability and dynamics.

Received 5 July; accepted 11 September 2007.



**Figure 4 | Probability that reefs of given initial state will remain above the unstable equilibrium during a 25-year period.** The physical disturbance regime includes stochastic hurricanes with a 20-year periodicity and the algal–coral overgrowth rate is  $8\text{ cm}^2\text{ yr}^{-1}$ . The unstable equilibrium is denoted with a thick dashed white line.

- Hughes, T. P. *et al.* Climate change, human impacts, and the resilience of coral reefs. *Science* **301**, 929–933 (2003).
- Connell, J. H. Disturbance and recovery of coral assemblages. *Coral Reefs* **16**, S101–S113 (1997).
- Pandolfi, J. M. *et al.* Global trajectories of the long-term decline of coral reef ecosystems. *Science* **301**, 955–958 (2003).
- Lessios, H. A. Mass mortality of *Diadema antillarum* in the Caribbean: What have we learned? *Annu. Rev. Ecol. Syst.* **19**, 371–393 (1988).
- Aronson, R. B. & Precht, W. F. White-band disease and the changing face of Caribbean coral reefs. *Hydrobiologia* **460**, 25–38 (2001).
- Knowlton, N. Thresholds and multiple stable states in coral reef community dynamics. *Am. Zool.* **32**, 674–682 (1992).
- Ludwig, D., Walker, B. & Holling, C. S. Sustainability, stability and resilience. *Conserv. Ecol.* **1**, 1–27 (1997).
- May, R. M. Thresholds and breakpoints in ecosystems with a multiplicity of stable states. *Nature* **269**, 471–477 (1977).
- Gardner, T. A., Cote, I. M., Gill, J. A., Grant, A. & Watkinson, A. R. Long-term region-wide declines in Caribbean corals. *Science* **301**, 958–960 (2003).
- Hughes, T. P. Catastrophes, phase shifts, and large-scale degradation of a Caribbean coral reef. *Science* **265**, 1547–1551 (1994).
- McClanahan, T. R. & Muthiga, N. A. An ecological shift in a remote coral atoll of Belize over 25 years. *Environ. Conserv.* **25**, 122–130 (1998).
- Petraitis, P. S. & Dudgeon, S. R. Detection of alternative stable states in marine communities. *J. Exp. Mar. Biol. Ecol.* **300**, 343–371 (2004).
- Mumby, P. J. The impact of exploiting grazers (Scaridae) on the dynamics of Caribbean coral reefs. *Ecol. Appl.* **16**, 747–769 (2006).
- Bythell, J. C., Gladfelter, E. H. & Bythell, M. Chronic and catastrophic natural mortality of three common Caribbean reef corals. *Coral Reefs* **12**, 143–152 (1993).
- de Ruiter van Steveninck, E. D. & Breeaman, A. M. Deep water populations of *Lobophora variegata* (Phaeophyceae) on the coral reef of Curaçao: influence of grazing and dispersal on distribution patterns. *Mar. Ecol. Prog. Ser.* **38**, 241–250 (1987).
- Meesters, E. H., Wesseling, I. & Bak, R. P. M. Coral colony tissue damage in six species of reef-building corals: partial mortality in relation with depth and surface area. *J. Sea Res.* **37**, 131–144 (1997).

17. Mumby, P. J., Foster, N. L. & Glynn Fahy, E. A. Patch dynamics of coral reef macroalgae under chronic and acute disturbance. *Coral Reefs* **24**, 681–692 (2005).
18. Box, S. J. & Mumby, P. J. The effect of macroalgal competition on the growth and survival of juvenile Caribbean corals. *Mar. Ecol. Prog. Ser.* **342**, 139–149 (2007).
19. Nugues, M. M. & Bak, R. P. M. Differential competitive abilities between Caribbean coral species and a brown alga: a year of experiments and a long-term perspective. *Mar. Ecol. Prog. Ser.* **315**, 75–86 (2006).
20. Mumby, P. J. *et al.* Fishing, trophic cascades, and the process of grazing on coral reefs. *Science* **311**, 98–101 (2006).
21. Murray, J. D. *Mathematical Biology 1: An Introduction* (Springer, New York, 2002).
22. Hoegh-Guldberg, O. Climate change, coral bleaching and the future of the world's coral reefs. *Mar. Freshwat. Res.* **50**, 839–866 (1999).
23. Carpenter, R. C. & Edmunds, P. J. Local and regional scale recovery of *Diadema* promotes recruitment of scleractinian corals. *Ecol. Lett.* **9**, 271–280 (2006).
24. Bellwood, D. R., Hughes, T. P., Folke, C. & Nystrom, M. Confronting the coral reef crisis. *Nature* **429**, 827–833 (2004).
25. Done, T. J. Coral community adaptability to environmental change at the scales of regions, reefs and reef zones. *Am. Zool.* **39**, 66–79 (1999).
26. Hildebrand, F. B. *Introduction to Numerical Analysis* (Dover Publications, New York, 1987).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This study was funded by grants from the Environmental Protection Agency, The Royal Society, the Natural Environment Research Council and the National Science Foundation. We thank P. Armsworth and H. Possingham for discussions and comments on the manuscript, respectively.

**Author Contributions** P.J.M. jointly conceived the study with A.H., designed and implemented the simulation model, and prepared the manuscript; A.H. created the analytic model with contributions from H.J.E. and P.J.M., supervised its analysis and edited the manuscript; H.J.E. analysed and described the analytic model, carried out the stochastic implementation of the simulation model, and edited the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to P.J.M. ([p.j.mumby@ex.ac.uk](mailto:p.j.mumby@ex.ac.uk)).

## LETTERS

## Neural mechanisms mediating optimism bias

Tali Sharot<sup>1,2</sup>, Alison M. Riccardi<sup>1</sup>, Candace M. Raio<sup>1</sup> & Elizabeth A. Phelps<sup>1</sup>

Humans expect positive events in the future even when there is no evidence to support such expectations. For example, people expect to live longer and be healthier than average<sup>1</sup>, they underestimate their likelihood of getting a divorce<sup>1</sup>, and overestimate their prospects for success on the job market<sup>2</sup>. We examined how the brain generates this pervasive optimism bias. Here we report that this tendency was related specifically to enhanced activation in the amygdala and in the rostral anterior cingulate cortex when imagining positive future events relative to negative ones, suggesting a key role for areas involved in monitoring emotional salience in mediating the optimism bias. These are the same regions that show irregularities in depression<sup>3</sup>, which has been related to pessimism<sup>4</sup>. Across individuals, activity in the rostral anterior cingulate cortex was correlated with trait optimism. The current study highlights how the brain may generate the tendency to engage in the projection of positive future events, suggesting that the effective integration and regulation of emotional and autobiographical information supports the projection of positive future events in healthy individuals, and is related to optimism.

People tend to make overly confident, positive predictions about the future, which are often inaccurate<sup>1,2</sup>. The tendency to expect good things in the future is known as optimism. Extreme optimism can be harmful as it can promote an underestimation of risk and poor planning<sup>5</sup>. In contrast, a pessimistic view is correlated with severity of depression symptoms<sup>3</sup>. A moderate optimistic illusion, however, can motivate adaptive behaviour in the present towards a future goal, and has been related to mental<sup>6</sup> and physical<sup>7</sup> health. How does the healthy brain generate the tendency to create images of positive future events?

It has been suggested that imagining future events demands a system that calls on the past to retrieve pieces of information that are then reconstructed to form representations of possible future scenarios<sup>8,9</sup>. In accordance with this proposal, it has been shown that imagining the future depends on the same neural networks that are active when recalling the past<sup>8,10,11</sup>. One possibility is that in optimists this network is more active when imagining future positive events than negative ones. It is also possible that other brain areas modulate this network in a way that biases it to engage in positive future projections.

To examine the neurobiological basis of optimism we collected functional magnetic resonance imaging (fMRI) data while participants thought of autobiographical events related to a description of a life episode (for example, 'winning an award' or 'the end of a romantic relationship'). The word 'past' or 'future' indicated if they should think of an event that occurred in the past or one that might occur in the future. Trials were classified into positive, negative and neutral according to participants' ratings. The mean number of trials rated as neutral was negligible, and thus all contrasts of interest were conducted between positive and negative trials only.

After scanning, participants rated their memories and projections on six factors<sup>12</sup> related to their subjective experience. The mean scores for these ratings and statistical analysis are presented in Table 1 and Supplementary Table 1. Finally, participants completed the LOT-R (Life Orientation Test-Revised) scale that measures trait optimism<sup>13</sup> (see Methods for details).

As represented in Table 1 and Supplementary Table 1, future positive events were rated as more positive than past positive events, and were imagined to be closer in temporal proximity than future negative events and all past events (Fig. 1a). Negative future events were experienced with a weaker subjective sense of pre-experiencing, and were more likely to be imagined from an outsider viewing in, than positive future events and all past events (Fig. 1b). The more optimistic participants were, as indicated by the LOT-R scores, the more likely they were to expect positive events to happen closer in the future than negative events, and to experience them with a greater sense of pre-experiencing (Fig. 1c, d).

Next we examined the fMRI data to identify the underlying neural mechanisms. Functional regions of interest (ROIs) were identified by contrasting all trials with fixation ( $P < 0.00005$ , uncorrected). The four regions that survived this threshold (Fig. 2a, b) included the rostral anterior cingulate cortex (rACC, Fig. 2b), extending into the ventral medial prefrontal cortex at a more lenient threshold (see Fig. 3a, c), the posterior cingulate cortex (PCC, Fig. 2b), and the dorsal medial prefrontal cortex (Fig. 2b), all of which have been reported to have a key role in retrieval of autobiographical memory and in imagining future events<sup>8</sup>. The fourth region was the right amygdala, which has been shown to be important for emotion's

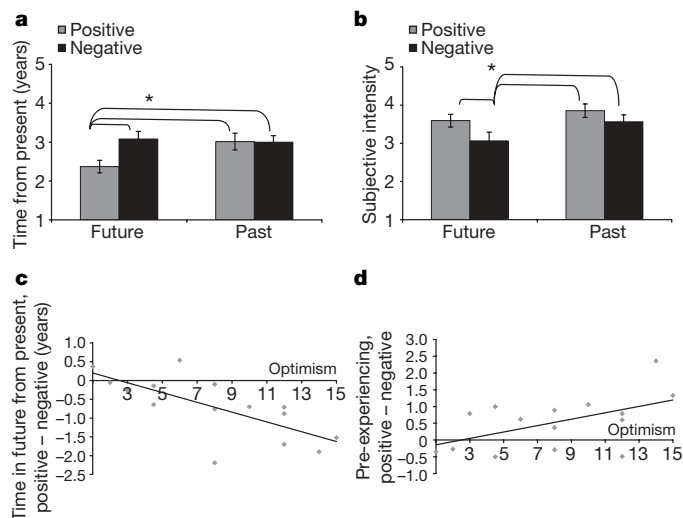
**Table 1 | Means and statistics for subjective ratings according to trial type**

Scale	Future				Past			
	Positive	Negative	t-test (P)	Correlation with optimism	Positive	Negative	t-test (P)	Correlation with optimism
Valence	2.44	4.34	<0.001	NS	2.77	4.07	<0.01	NS
Arousal	3.56	3.23	= 0.05	NS	3.50	3.61	NS	NS
Time of event	2.37	3.08	<0.005	<0.005	3.01	3.00	NS	NS
Subjective temporal distance	3.00	3.47	NS	NS	3.33	3.40	NS	NS
Perspective	2.96	3.37	<0.025	NS	2.92	2.91	NS	NS
Reliving or pre-experiencing	3.59	3.07	<0.025	<0.025	3.80	3.85	NS	NS
Vividness	3.66	3.28	NS	NS	3.90	4.19	NS	NS

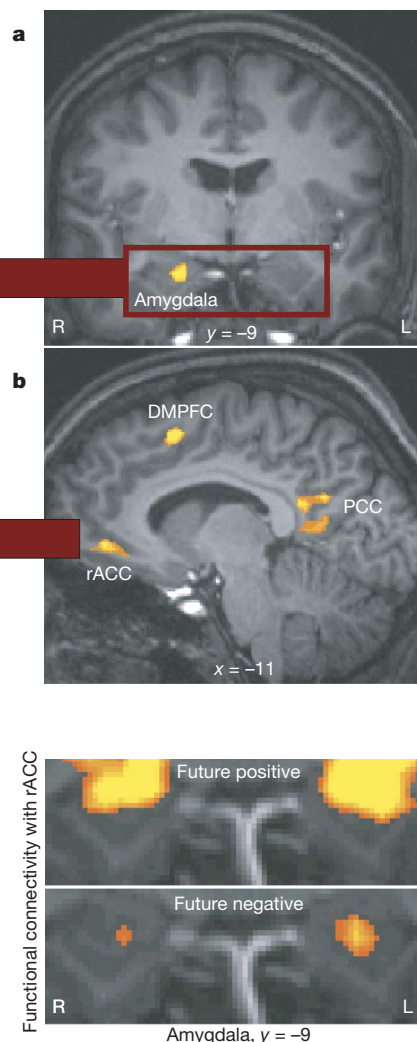
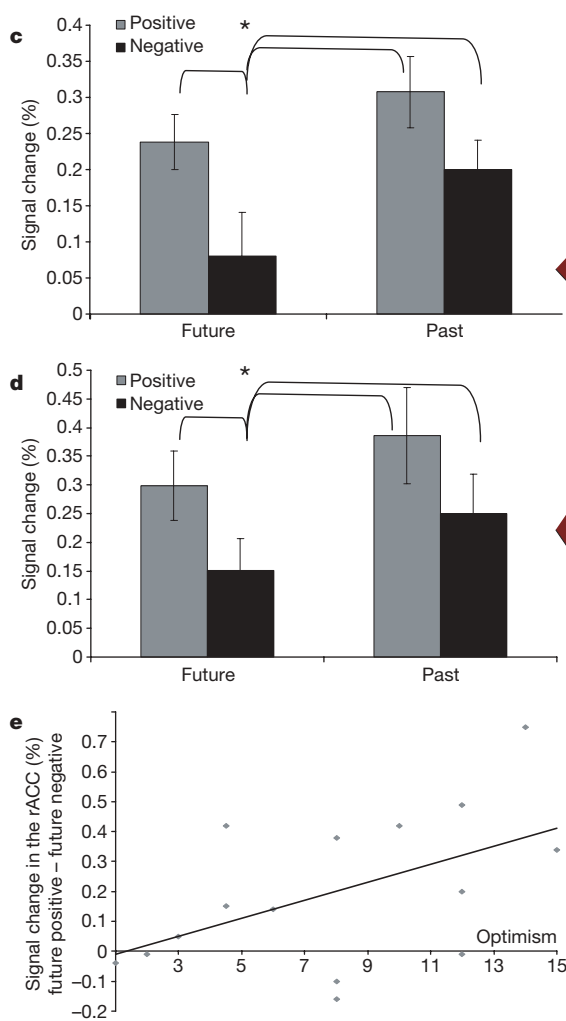
Valence (1, positive; 6, negative), arousal (1, low; 6, high), time of event (years from present, 1–5), subjective temporal distance (1, near; 6, far), perspective (1, involved; 6, observer), reliving or pre-experiencing (1, low; 6, high) and vividness (1, low; 6, high). NS, not significant.

<sup>1</sup>Department of Psychology, New York University, New York 10003, USA. <sup>2</sup>Wellcome Department of Imaging Neuroscience, Institute of Neurology, University College London, London WC1N 3BG, UK.





**Figure 1 | Optimism related to expected time of event and sense of pre-experiencing future events.** **a**, Positive events are perceived to be closer in time to the present than negative future events and all past events. **b**, Negative future events are experienced with a weaker sense of pre-experiencing than positive future events and all past events. **c**, The difference in expected time of event for future positive and negative events correlated with the optimism level defined by ranking participants according to their LOT-R scores ( $r = -0.74$ ,  $P < 0.005$ ). **d**, The difference in pre-experiencing for positive and negative events correlated with trait optimism ( $r = 0.58$ ,  $P < 0.025$ ).  $N = 15$ ; error bars,  $\pm$ s.e.m.; \* $P < 0.05$ .

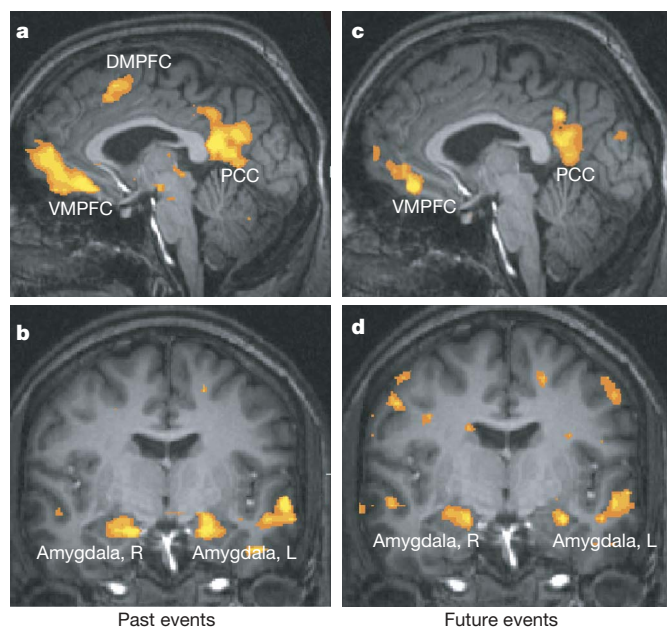


**Figure 2 | Activity in the amygdala and rACC and its relation to optimism.** ROIs defined by contrasting all trials with fixation (shown at  $P < 0.0005$ , uncorrected) reveal activity of **a**, the amygdala (Talarich coordinates of peak voxel: 20, -9, -14) and **b**, the rACC (Brodmann Area (BA) 32: -11, 42, -1), the PCC (BAs 30 and 31: -4, -47, 19) and the dorsal medial prefrontal cortex (DMPFC; BA 6: -8, 12, 52). BOLD signals in both **c**, the amygdala and **d**, the rACC reduced while imagining negative future events. **e**, BOLD signal in the rACC during future positive events versus future negative events correlated with trait optimism ( $-9, 39, -2$ ;  $r = 0.5$ ,  $P < 0.05$ ). **f**, Functional connectivity map with rACC as the seed reveals extensive correlation with the amygdala signal during future positive trials, and restricted correlation during future negative trials. The red square in **a** indicates the region shown in **f**.  $N = 15$ ; error bars,  $\pm$ s.e.m.; \* $P < 0.05$ .

influence on autobiographical memory<sup>14</sup> (Fig. 2a; the left amygdala was also observed at a more lenient threshold, Fig. 3b, d). Among these ROIs, the blood-oxygenation-level-dependent (BOLD) signal in the amygdala and rACC was reduced when imagining negative future events relative to positive future events and to all past events (Fig. 2c, d). The differences in BOLD signal between positive and negative future events remained significant when controlling for differences in the sense of pre-experiencing (amygdala,  $P < 0.01$ ; rACC,  $P < 0.05$ ). For structural ROI analysis (Supplementary Fig. 1) and exploratory whole-brain analysis, see Supplementary Information.

To examine if these differences were related to optimism across individuals, we identified voxels within the four functional ROIs in which changes in BOLD signal during future positive trials relative to future negative trials were correlated with participants' optimism score on the LOT-R scale (see Supplementary Information for additional analysis and discussion). We found a positive correlation in the rACC (Fig. 2e). No significant correlation with optimism was observed in the other ROIs. Finally, a functional connectivity analysis revealed a strong correlation between activity in the rACC and activity in the amygdala bilaterally while imagining future positive events; this correlation was weaker and less extensive when imagining future negative events (Fig. 2f).

Our behavioural results suggest that, whereas the past is constrained, the future is open to interpretation allowing subjects to distance themselves from possible negative events and move closer toward positive ones. Across individuals, this tendency was associated with trait optimism. The brain imaging findings provide a possible mechanism mediating the behavioural observations.



**Figure 3 | Brain activity during past and future trials.** Regions that emerged when contrasting **a, b**, past trials with fixation and **c, d**, future trials with fixation at a threshold of  $P < 0.005$ , uncorrected, reveal engagement of regions shown previously<sup>8</sup> to be involved in autobiographical memory and future projection including ventral medial prefrontal cortex (VMPFC) and the PCC (**a, c**), as well as the amygdala bilaterally (**b, d**).

Reduced BOLD signal was observed in the amygdala and rACC during imagination of negative future events relative to positive future events and all past events, suggesting that the optimism bias may be related to a reduction in negative future thought.

The amygdala has a documented role in the modulation by emotion of cognitive processes including memory and decision making (for a review, see ref. 15). Our data extend the role of the amygdala to include simulation of future emotional events. Previous studies examining the neural mechanisms underlying imagination of non-emotional future events did not observe amygdala involvement<sup>8,10</sup>. Thus, as in the case of memory, we suggest that the amygdala may be selectively engaged in imagining future emotional events, rather than serving a general function in imagining the future.

The rACC has strong reciprocal connection with the amygdala and other regions that convey emotional and motivational information<sup>16</sup>. It has been implicated in tasks involving self-reflection, such as making positive self-referential judgments<sup>17</sup>, reflecting on hopes and dreams<sup>18</sup>, indicating preferences<sup>19</sup> and judging the trustworthiness of others<sup>20</sup>. This has caused some researchers to suggest that when information is judged to be self-relevant, activity in the rACC will convey the valence<sup>17</sup>. However, others have suggested a more general role for the rACC in assessing the salience of emotional and motivational information and regulating emotional responses accordingly<sup>21</sup>. Consistent with this notion it has been shown that the rACC's response to positive and negative stimuli reflects situational specificity. For example, both the rACC and amygdala were more sensitive to positive stimuli in subjects that focused on obtaining goals (promotional context) and to negative stimuli in subjects that focused on avoiding failure (prevention context)<sup>22</sup>. Furthermore, changes in rACC activity have been related to extinction of fear conditioning<sup>23</sup>, and to reduced anxiety during shock expectancy<sup>24</sup>; this possibly indicates diminished salience of negative stimuli that may consequently lower amygdala activity<sup>23</sup>.

Consistent with these previous studies, we suggest that in the current study the rACC is tracking the subjective salience of the stimuli by assessing emotional, motivational and autobiographical information, and possibly by regulating such signals. This suggestion is

supported by the strong functional connectivity observed between the rACC and amygdala during imagination of positive future events. In addition, individual differences in trait optimism were related to the relative level of rACC engagement when imagining positive future scenarios versus negative ones. We speculate that rACC activity during imagination of future events reflects a self-regulatory focus that underlies a bias in attention and vigilance towards positive future events and away from negative ones, modulating engagement of other regions that provide emotional and autobiographical information that is processed and recombined to construct future scenarios.

These findings may provide insight to the mechanisms underlying depression. Depressive symptoms are associated with pessimism<sup>4</sup> and with difficulties in creating detailed images of future events<sup>25</sup>. It has been suggested that malfunction of a neural pathway incorporating the rACC and the amygdala may cause depression by leading to decreased regulatory affects of the rACC over the amygdala and other regions involved in emotional processing<sup>3</sup>. Future studies are needed to determine whether these abnormalities are related directly to the breakdown of optimism in depression, specifically because it relates to the way in which depressed patients imagine future events.

One should note that when comparing past and future trials we are also comparing imagining and remembering. Thus, our data cannot indicate whether the positivity bias is a function of time (that is, it will emerge only when thinking about the future) or whether it reflects a tendency to engage in positive thought when not constrained by reality. We speculate that the difference is not between past and future, although in practice the optimism bias manifests itself in thoughts of the future because the absence of factual constraints leaves plenty of room to alter expectations.

Expecting positive events, and generating compelling mental images of such events, may serve an adaptive function by motivating behaviour in the present towards a future goal. We do not suggest that simulating negative events is not important for survival. However, pondering on such events may interfere with daily activities by promoting negative effects such as anxiety and depression. The current study highlights how the brain may generate the tendency to engage in the projection of positive future events, suggesting that the rACC modulates activity in brain areas that are involved in emotional processing and autobiographical retrieval to create positive images of the future.

## METHODS SUMMARY

Functional MRI images were collected from participants while they thought of autobiographical events related to a description of a life episode that appeared on screen for 14 s. Either the word 'past' or 'future' indicated if they should think of an event that occurred in the past or that might occur in the future. The participants were instructed to press a button once the memory or projection was beginning to form in their mind, and again when they finished elaborating on it. They then had 2 s to rate the memory or projection for emotional arousal and 2 s to rate it for valence. There were 80 trials that were classified into positive, negative or neutral according to the participants' ratings.

After scanning, participants rated their memories or projections on six factors<sup>12</sup> related to their subjective experience, and completed the LOT-R scale that measures trait optimism<sup>13</sup>. Participants were ranked according to their optimism score on the LOT-R scale. These optimism rankings were correlated with the difference in scores (future positive minus future negative) on the behavioural scales.

For each participant a time series was created indicating the temporal position of the different trial types triggered from the participant's first button press. Functional ROIs were identified by contrasting all trials with fixation ( $P < 0.00005$ , uncorrected) using random effects GLM (General Linear Model) on the data of all participants. The time courses of activation were extracted from these regions for each trial type and participant for statistical analysis. Next, we identified voxels within the ROIs in which changes in BOLD signal during future positive trials relative to future negative trials were correlated with the LOT-R scores. Additionally, functional connectivity maps<sup>26</sup> were created for the different trial types (False Discovery Rate (FDR) correction threshold at  $t = 5.5$ ).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 7 July; accepted 19 September 2007.**

**Published online 24 October 2007.**

- Weinstein, N. D. Unrealistic optimism about future life events. *J. Pers. Soc. Psychol.* **39**, 806–820 (1980).
- Hoch, S. Counterfactual reasoning and accuracy in predicting personal events. *J. Exp. Psychol.* **11**, 719–731 (1984).
- Drevets, W. C. *et al.* Subgenual prefrontal cortex abnormalities in mood disorders. *Nature* **386**, 824–827 (1997).
- Strunk, D. R., Lopez, H. & DeRubeis, R. J. Depressive symptoms are associated with unrealistic negative predictions of future life events. *Behav. Res. Ther.* **44**, 861–882 (2006).
- Lovallo, D. & Kahneman, D. Delusions of success: how optimism undermines executives' decisions. *Harv. Bus. Rev.* **56**, 56–63 (2003).
- Taylor, S. E. & Brown, J. D. Illusion and well-being: a social psychological perspective on mental health. *Psychol. Bull.* **103**, 193–210 (1988).
- Scheier, M. F. & Carver, C. S. Dispositional optimism and physical well-being: the influence of generalized outcome expectancies on health. *J. Pers.* **55**, 169–210 (1987).
- Addis, D. R., Wong, A. T. & Schacter, D. L. Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. *Neuropsychologia* **45**, 1363–1377 (2007).
- Tulving, E. in *The Missing Link in Cognition* (eds Terrace, H. S. & Metcalfe, J.) 3–56 (Oxford Univ. Press, New York, 2005).
- Okuda, J. *et al.* Thinking of the past and past: the roles of the frontal pole and the medial temporal lobe. *Neuroimage* **19**, 1369–1380 (2003).
- Hassabis, D., Kumaran, D., Vann, S. D. & Marguire, E. A. Patients with hippocampal amnesia cannot imagine new experiences. *Proc. Natl Acad. Sci. USA* **104**, 1726–1731 (2007).
- D'Argembeau, A. & Van der Linden, M. Influence of affective meaning on memory for contextual information. *Emotion* **4**, 173–188 (2004).
- Scheier, M. F., Carver, C. S. & Bridges, M. W. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. *J. Pers. Soc. Psychol.* **67**, 1063–1078 (1994).
- Sharot, T., Martorella, E. A., Delgado, M. R. & Phelps, E. A. How personal experience modulates the neural circuitry of memories of September 11. *Proc. Natl Acad. Sci. USA* **104**, 389–394 (2007).
- Phelps, E. A. Emotion and cognition: insights from studies of the human amygdala. *Annu. Rev. Psychol.* **57**, 27–53 (2006).
- Vogt, B. A. & Pandya, D. N. Cingulate cortex of the rhesus monkey: II. cortical afferents. *J. Comp. Neurol.* **262**, 271–289 (1987).
- Moran, J. M., Macrae, C. N., Heatherton, T. F., Wyland, C. L. & Kelley, W. M. Neuroanatomical evidence for distinct cognitive and affective components of self. *J. Cogn. Neurosci.* **18**, 1586–1594 (2006).
- Johnson, S. C., Baxter, L. C., Wilder, L. S., Pipe, J. G., Heiserman, J. E. & Prigatano, G. P. Neural correlates of self-reflection. *Brain* **125**, 1808–1814 (2002).
- Paulus, M. P. & Frank, L. R. Ventromedial prefrontal cortex activation is critical for preference judgments. *Neuroreport* **14**, 1311–1315 (2003).
- Winston, J. S., Strange, B. A., O'Doherty, J. & Dolan, R. J. Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neurosci.* **5**, 277–283 (2002).
- Bush, G., Luu, P. & Posner, M. I. Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn. Sci.* **14**, 215–222 (2000).
- Cunningham, W. A., Raye, C. L. & Johnson, M. K. Neural correlates of evaluation associated with promotion and prevention regulatory focus. *Cogn. Affect. Behav. Neurosci.* **5**, 202–211 (2005).
- Phelps, E. A., Delgado, M. R., Nearing, K. I. & LeDoux, J. E. Extinction learning in humans: role of the amygdala and vmPFC. *Neuron* **43**, 897–905 (2004).
- Simpson, J. R., Snyder, A. Z. Jr, Gusnard, D. A. & Raichle, M. E. Emotion-induced changes in human medial prefrontal cortex: I. during cognitive task performance. *Proc. Natl Acad. Sci. USA* **98**, 688–693 (2001).
- Williams, J. M., Ellis, N. C., Tyers, C., Healy, H., Rose, G. & MacLeod, A. K. The specificity of autobiographical memory and imageability of the future. *Mem. Cognit.* **24**, 116–125 (1996).
- Roebroek, A., Formisano, E. & Goebel, R. Mapping directed influence over the brain using Granger causality and fMRI. *Neuroimage* **25**, 230–242 (2005).
- Pruessner, J. C. *et al.* Volumetry of hippocampus and amygdala with high-resolution MRI and three-dimensional analysis software: minimizing the discrepancies between laboratories. *Cereb. Cortex* **10**, 433–442 (2000).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This study was supported by the NIMH (E.A.P.), the Seaver Foundation (grant to NYU's Center for Brain Imaging), and a Margaret and Herman Sokol Postdoctoral Fellowship (T.S.). We thank S. H. Mulhern, D. C. Johnson and J. K. Szary for help in data analysis, and D. Schiller, J. H. McDermott and Y. Trope for discussion.

**Author Contributions** T.S. designed the study. T.S. and E.A.P. interpreted the data and wrote the paper. T.S. and A.M.R. developed stimuli, gathered behavioural pilot data, and conducted behavioural data analysis. T.S. gathered fMRI data. T.S. conducted neuroimaging analyses with the help of A.M.R. and C.M.R., and with advice of E.A.P. Whole-brain exploratory analysis was conducted by C.M.R.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to E.A.P. ([liz.phelps@nyu.edu](mailto:liz.phelps@nyu.edu)).



## METHODS

**Participants.** Participants (18) were recruited through posted advertisements. Three were eliminated from the analysis because of either excessive head movement or spikes identified in the fMRI data. The remaining 15 participants were included in the analysis (7 males, 8 females; mean age, 23.4 yr; range, 18–36 yr). Before scanning, participants completed a screening form for significant medical conditions. All participants gave informed consent and were paid for their participation.

**Stimuli.** Stimuli consisted of 80 short descriptions of possible life episodes (Supplementary Information). The order and condition in which episodes were presented were random. Because participants had a tendency to generate positive events, we needed to provide more negative and neutral life episodes to obtain an equal number of trials rated as positive and negative.

**Procedure.** Before the scanning session, participants went through four practice trials. The session began with a short structural scan, followed by four functional scans (10 min each), each consisting of 20 trials of 30 s. Finally, an additional longer structural scan was performed. Stimuli were presented by means of a mirror mounted on the head coil. In each trial, a description of a possible life episode and the word indicating the condition of the trial (either 'past' or 'future') appeared for 14 s. The participants were instructed to press a button once a mental memory or projection was beginning to form in their mind, and again if they finished elaborating on it before the end of the trial. The participant then had 2 s to rate the memory or projection for emotional arousal (1, not arousing; 2, somewhat arousing; 3, highly arousing) and 2 s to rate it for valence (1, positive; 2, neutral; 3, negative). Finally, a fixation cross was presented for 12 s. If the participant did not press the button to indicate that they were thinking of a specific memory or projection, or did not indicate arousal and/or valence by pressing the button box, that trial was excluded from the final data analysis. The number of trials rated as positive and negative did not differ for future or past conditions. Reaction times, indicating when a mental memory or projection was beginning to form in the participants' mind, did not differ for the different trial types (range of means: 2.9–3.6 s).

After the scanning session, participants were presented with the same trials again on a computer screen, and were asked to rate their memories or projections on six scales adapted from a previous study<sup>12</sup>: vividness, sense of reliving or pre-experiencing, estimation of the time of the event, subjective temporal distance from the event, perspective (involved or observer), arousal and valence. Finally, participants completed the LOT-R scale that measures trait optimism–pessimism on a scale from 0 (pessimistic) to 24 (optimistic)<sup>13</sup>. Participants were generally optimistic; the mean optimism score was 17.76.

**Behavioural analysis.** Participants were ranked according to their optimism score on the LOT-R scale from the most optimistic to the least optimistic. These optimism rankings were correlated with the difference scores (future positive minus future negative) on the behavioural scales.

**MRI scanning and data analysis.** The study was conducted at the NYU Center for Brain Imaging using a 3T Siemens Allegra scanner and a Siemens head coil. Anatomical images were acquired using MPRage scans, followed by 3-mm-thick axial slices (parallel to the anterior cingulate–posterior cingulate plane).

Positioning of the slices was based on an AAScout sequence. Functional scans used a gradient echo sequence: TR (time of repetition) = 2 s, TE (time of echo) = 25 ms, FA (flip angle) = 90, matrix =  $64 \times 64$ , FOV (field of view) = 192 mm, slice thickness = 3 mm. A total of 39 axial slices parallel to the AC–PC plane were sampled for whole-brain coverage. The in-plane resolution was 3 mm  $\times$  3 mm. Imaging data were analysed using Brain Voyager software (QX, version 1.7).

Data were temporally and spatially smoothed (4 mm full-width at half-maximum) and motion-corrected. Individual data were transformed into Talairach space for group analysis. For each participant, a time series was created indicating the temporal position of the different trial types triggered from the participant's first button press to the onset of the rating scale.

Trial types were classified into six groups according to the participants' arousal and valence ratings during the fMRI scanning: future positive, future negative, future neutral, past positive, past negative and past neutral. Negative events were those receiving high (3) or medium (2) arousal rating and negative valence (3). Positive events were those receiving high (3) or medium (2) arousal rating and positive valence (1). Neutral events were those receiving low arousal rating (1) and neutral valence (2). Trials receiving any other combination of responses could not be characterized in either group and were not included in the analysis. The seventh trial type was fixation.

Functional ROIs were identified by contrasting all trials with fixation ( $P < 0.00005$ , uncorrected) using random effects GLM on the data of all participants. We extracted the time course of activation from these regions for the different trial types for each participant. The average percentage signal change from 6 s to 8 s after the participant's first button press in these ROIs were calculated for each trial type and participant. Statistical tests were performed on these values. Within these functional ROIs, we conducted a voxel-wise analysis of covariance contrasting future positive trials with future negative trials using the LOT-R optimism score as a covariate ( $P < 0.05$ ). In addition, a priori regions of interest—the right and left amygdala—were identified according to landmarks described previously<sup>27</sup>. Extraction of time courses and statistical analysis was done in the same manner as for the functional ROIs.

Functional connectivity analysis<sup>26</sup> was conducted using the rACC ROI as a seed. Maps were created to identify voxels where BOLD signal was temporally correlated with the seed during future positive trials and during future negative trials (FDR correction threshold at  $t = 5.5$ ).

To identify other voxels in the brain that showed stronger BOLD responses during one trial type than the other, we conducted a whole-brain exploratory analysis on group data using a random effects GLM (at  $P < 0.001$ , uncorrected;  $> 10$  contiguous voxels 1 mm<sup>3</sup>).

## LETTERS

# PYY modulation of cortical and hypothalamic brain areas predicts feeding behaviour in humans

Rachel L. Batterham<sup>1</sup>, Dominic H. ffytche<sup>2</sup>, J. Miranda Rosenthal<sup>1</sup>, Fernando O. Zelaya<sup>2</sup>, Gareth J. Barker<sup>2</sup>, Dominic J. Withers<sup>1</sup> & Steven C. R. Williams<sup>2</sup>

The ability to maintain adequate nutrient intake is critical for survival. Complex interrelated neuronal circuits have developed in the mammalian brain to regulate many aspects of feeding behaviour, from food-seeking to meal termination. The hypothalamus and brainstem are thought to be the principal homeostatic brain areas responsible for regulating body weight<sup>1,2</sup>. However, in the current 'obesogenic' human environment food intake is largely determined by non-homeostatic factors including cognition, emotion and reward, which are primarily processed in corticolimbic and higher cortical brain regions<sup>3</sup>. Although the pleasure of eating is modulated by satiety and food deprivation increases the reward value of food, there is currently no adequate neurobiological account of this interaction between homeostatic and higher centres in the regulation of food intake in humans<sup>1,4,5</sup>. Here we show, using functional magnetic resonance imaging, that peptide YY<sub>3–36</sub> (PYY), a physiological gut-derived satiety signal, modulates neural activity within both corticolimbic and higher-cortical areas as well as homeostatic brain regions. Under conditions of high plasma PYY concentrations, mimicking the fed state, changes in neural activity within the caudolateral orbital frontal cortex predict feeding behaviour independently of meal-related sensory experiences. In contrast, in conditions of low levels of PYY, hypothalamic activation predicts food intake. Thus, the presence of a postprandial satiety factor switches food intake regulation from a homeostatic to a hedonic, corticolimbic area. Our studies give insights into the neural networks in humans that respond to a specific satiety signal to regulate food intake. An increased understanding of how such homeostatic and higher brain functions are integrated may pave the way for the development of new treatment strategies for obesity.

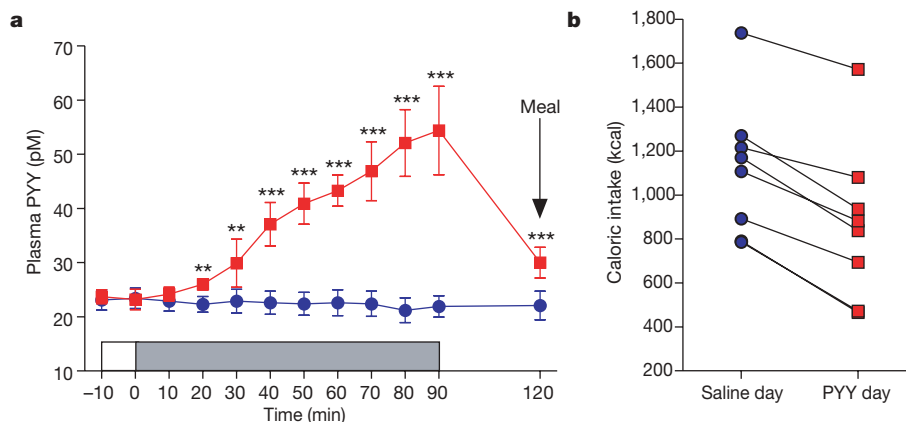
Energy homeostasis is controlled by long-term adiposity hormones and short-term gastrointestinal signals, reflecting overall energy stores and individual meal intake, respectively<sup>1,2</sup>. Rodent studies have shown that the brainstem and hypothalamus serve as key homeostatic sites receiving and integrating neural, nutrient and hormonal signals and orchestrating appropriate efferent responses<sup>1,2</sup>. However, the rewarding aspects of food, environmental cues, and other emotional and cognitive factors have an important role in regulating feeding behaviour in humans<sup>3</sup>. Previous functional imaging studies have investigated brain activity in the fasted state and mapped the brain regions that respond to the ingestion of nutrients such as glucose, chocolate or a liquid meal<sup>6–10</sup>. However, these experiments have been complicated by the fact that eating per se elicits complex multi-sensory outputs including olfactory, gustatory, cognitive and emotional responses. Feeding also causes gut distension, vagal activation, and alters circulating nutrients and hormones. Furthermore, these studies did not link brain activity levels to

subsequent food intake and have therefore given limited insights into the circuits regulating feeding behaviour in humans. Anatomical and temporal constraints in some of these studies also mean that the whole neural network involved in such processes has not been modelled.

Gut hormones have key roles in the regulation of energy homeostasis<sup>11,12</sup>. For example, infusion of PYY, which is released from gut L-cells postprandially, reduces food intake in humans<sup>13,14</sup>. In contrast, plasma levels of ghrelin, a stomach-derived hormone, increase in the fasted state and infusion of this hormone stimulates food intake<sup>15</sup>. Gut hormones therefore represent physiological tools for probing both the homeostatic and higher neuroanatomical substrates that potentially predict feeding behaviour and signal satiety in man. We developed a protocol combining intravenous administration of PYY with whole-brain blood-oxygen-level-dependent (BOLD) functional magnetic resonance imaging (fMRI) and physiological and behavioural measures including assessment of subsequent food intake. Eight right-handed, normal-weight, food-deprived subjects participated in a randomized, double-blind, placebo-controlled crossover study (Supplementary Fig. 1 and Methods). Each subject underwent a 100-min continuous fMRI scanning protocol, consisting of a 10-min baseline scanning period while saline was infused intravenously, followed by a 90-min scanning period, during which subjects received either saline or PYY. Visual analogue scales (VAS) presented at minute intervals assessed the effects of PYY on food- and non-food-related subjective feelings (Supplementary Fig. 1 and Methods). Blood was sampled every 10 min for measurement of plasma hormones and glucose. Subjects were blinded to the study condition and during the infusion/scanning period were devoid of sensory food-related cues. This permitted investigation of circuits modulated by PYY independently of the environmental and neurological changes engendered by food-related sensory stimuli. To correlate changes in brain activity with feeding behaviour, we measured the subjects' caloric intake during a test meal 30 min after the end of the scanning period.

Circulating levels of PYY are low and progressively fall in the fasted state<sup>16</sup>. After eating, plasma PYY levels increase in proportion to the calories ingested<sup>17</sup>, signalling nutrient ingestion to appetite regulating circuits and reducing subsequent food intake<sup>13,14</sup>. The PYY study limb therefore examined the effects of a physiological satiety factor on brain activity and food intake, in part mimicking the endocrine environment of the fed state, whereas the saline infusion day examined brain activity in the fasted, low-PYY state. PYY infusion achieved plasma concentrations equivalent to those normally seen postprandially<sup>18</sup> (Fig. 1a), suggesting that the neural, humoral and behavioural changes we observed were physiologically relevant. During the scanning period there were no significant differences

<sup>1</sup>Centre for Diabetes and Endocrinology, Department of Medicine, University College London, London WC1E 6JJ, UK. <sup>2</sup>Centre for Neuroimaging Sciences, King's College London, Institute of Psychiatry, De Crespigny Park, Denmark Hill, London SE5 8AF, UK.



**Figure 1 | Effect of infusions on plasma PYY concentrations and food intake.** **a**, Plasma PYY concentrations during saline (blue circles) and PYY infusion (red squares) study days. The open box represents baseline 10-min saline infusion; the grey box indicates subsequent 90-min saline or PYY infusion. A meal was served at 120 min as indicated by the black arrow. Data

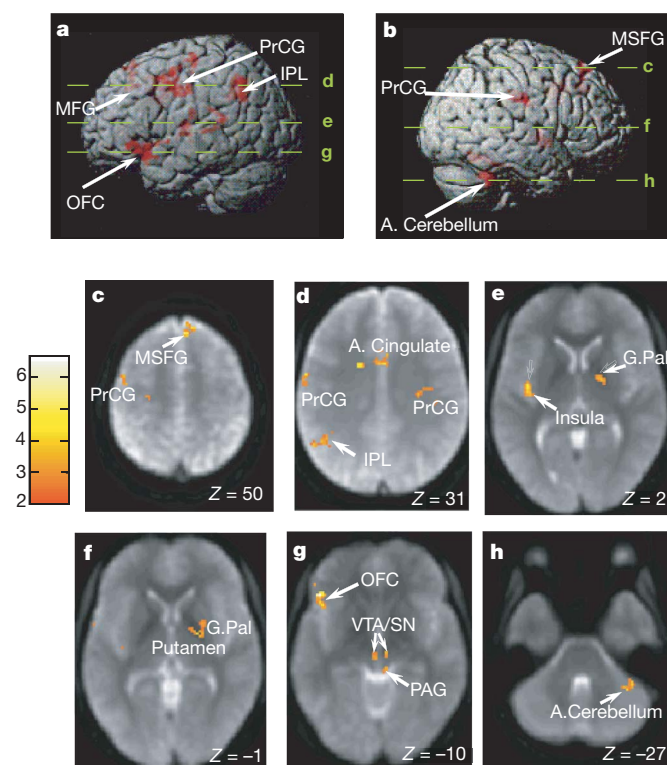
are mean  $\pm$  s.e.m. ( $n = 8$  per group). Double asterisk,  $P < 0.01$ ; triple asterisk,  $P < 0.001$ . **b**, Caloric intake from buffet meal 30 min after termination of saline or PYY infusion. Lines indicate individual changes in caloric intake for each subject between saline day (blue circle) and PYY day (red square).

between the subjective food-related or non-food-related VAS scores reported by the subjects on the PYY and saline infusion days. However, during the 30-min period before the meal, ratings of prospective food consumption significantly increased on the saline but not the PYY day (Supplementary Fig. 2a). PYY did not affect non-food-related VAS scores and elicited no negative subjective feelings. The caloric intake of all subjects was reduced by PYY by an average of  $24.8 \pm 4.1\%$ ,  $P = 0.00005$  (Fig. 1b) and was negatively correlated with the area under the curve (AUC) for the plasma PYY concentration time profile (Supplementary Fig. 2b). Despite consuming fewer calories on the PYY day, subjects rated their meal as equally pleasant on both study days (Supplementary Fig. 2c). PYY infusion caused a  $20.1 \pm 4.5\%$  reduction in active (acylated) ghrelin AUC compared with the saline infusion (Supplementary Fig. 2d), but did not alter plasma concentrations of glucose, insulin or leptin (Supplementary Table 1).

Rodent studies suggest that PYY acts on brainstem and hypothalamic circuits to exert its anorectic effects<sup>13,19</sup>. Recently, the midbrain ventral tegmental area (VTA) and nucleus accumbens (NAc) have been shown to respond to leptin and ghrelin, suggesting that these regions may also be involved in the integration of homeostatic signals<sup>20,21</sup>. We therefore hypothesized that PYY would modulate neural activity within the brainstem, midbrain (VTA/substantia nigra), hypothalamus and NAc (Supplementary Fig. 3).

We first examined brain regions that co-varied positively with plasma PYY concentrations (BOLD signal on both study days; increasing with PYY infusion on PYY study day and decreasing with the falling plasma PYY levels on the saline study day). After correction for multiple comparisons over the whole brain ( $P < 0.05$  cluster-level corrected), significant activation was seen in the brainstem parabrachial nucleus (PBN) and midbrain VTA (Fig. 2, Supplementary Fig. 4 and Supplementary Table 2). However, the greatest modulatory effect of PYY on brain activity was seen within the left caudolateral orbital frontal cortex (OFC), a polymodal brain region implicated in reward processing<sup>22,23</sup> (Fig. 2, Supplementary Figs 5–7 and Supplementary Table 2). Activation was also observed within the limbic system (insula and anterior cingulate cortices), ventral striatum (globus pallidus and putamen) and within discrete regions in the frontal, parietal, temporal and cerebellar cortices (Fig. 2 and Supplementary Table 2). BOLD signal in the right middle frontal gyrus and right angular gyrus co-varied negatively with plasma PYY levels, decreasing with PYY infusion and increasing with saline infusion (Supplementary Table 3). Although this whole-brain analysis revealed activity in some of our hypothesized brain regions, other regions were not activated or displayed asymmetrical activity,

prompting us to analyse further our specific a priori regions. Positive covariance with PYY was found within the posterior hypothalamus, right substantia nigra and right parabrachial nucleus at  $P < 0.05$ , voxel-corrected for the a priori region (Supplementary Figs 5, 8, 9



**Figure 2 | Brain areas modulated by PYY.** **a**, **b**, Surface-rendered views of the left (**a**) and right (**b**) hemispheres are displayed with superimposed group functional activity (threshold of  $P < 0.05$  cluster-level corrected). The image saturation (red) indicates the depth of the activation. Green dotted lines indicate levels of axial slices in **c–h**. **c–h**, Axial slices from group-averaged echo planar images with superimposed group functional activity (threshold of  $P < 0.05$  cluster-level corrected). The colour scale gives the  $T$  value of the functional activity. The left side of each slice is the left side of the brain.  $Z$  is the Talairach  $Z$  co-ordinate of the axial slice. Full coordinates and  $Z$  scores are presented in Supplementary Table 2. A, Cerebellum, anterior cerebellum; A. Cingulate, anterior cingulate; G.Pal, globus pallidus; IPL, inferior parietal lobule; MFG, middle frontal gyrus; MSFG, medial superior frontal gyrus; OFC, caudolateral orbitofrontal cortex; PAG, peri-aqueductal grey; PrCG, precentral gyrus; SN, substantia nigra; VTA, ventral tegmental area.



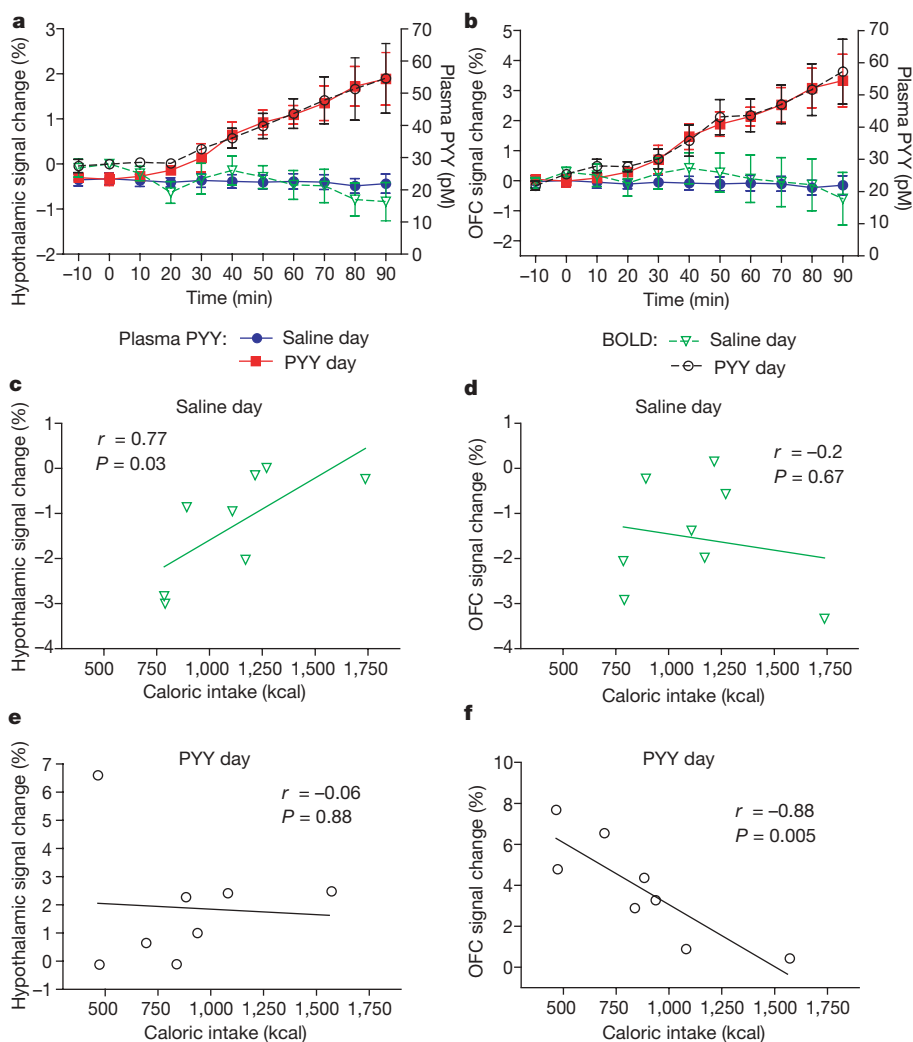
and Supplementary Table 4). However, no activation was observed within the NAc or solitary nucleus or tract at the same threshold. No brain regions co-varied with plasma ghrelin concentrations at the level of corrected significance used for the PYY analyses. At an uncorrected significance threshold, regions of covariance with ghrelin were observed within the hypothalamus, VTA and brainstem, regions implicated by rodent studies<sup>11,20</sup>.

To exclude whether responding to questions about internal state might account for changes in brain activity, we carried out an event-related analysis of VAS assessments. Food-related questions evoked brain activity in frontal, parietal, cingulate and temporal cortices with partial overlap with the cortical regions co-varying with PYY, in particular the pre-central gyrus and anterior cingulate cortex (Supplementary Fig. 10). However, these activations were not specific to food (no activations were seen for the food greater than non-food contrast at the same significance threshold), suggesting that they related to internal state appraisal and VAS response. Critically, no VAS-related activity was found within the caudolateral OFC, hypothalamus, brainstem, midbrain, ventral striatum, insula or the inferior parietal lobule, a higher cortical area co-varying with PYY (Supplementary Fig. 10).

Next we investigated whether changes in brain activity within the hypothalamus or OFC predicted subsequent feeding behaviour. We correlated the signal change extracted from the time series data from the PYY analysis for the hypothalamus and caudolateral OFC (Fig. 3a, b and Supplementary Figs 6–9) with caloric intake for each subject on the PYY and saline study days (Fig. 3c–f). On the saline infusion day, when the subjects were fasted and had low plasma PYY levels, the

change in signal within the hypothalamic region correlated with subsequent caloric intake, accounting for 59% of the variance (Fig. 3c). However, there was no correlation between signal change within the OFC and caloric intake (change in OFC signal predicted 4% of caloric intake variance, Fig. 3d). In contrast, on the PYY day, in the face of postprandial levels of PYY, change in OFC signal now predicted 77% of caloric intake variance whereas hypothalamic signal change accounted for less than 1% (Fig. 3e, f). Thus, in the presence of PYY, a postprandial satiety factor, brain activity predicting caloric intake appeared to switch from a homeostatic area (hypothalamus) to a hedonic area (OFC).

Our studies give the first insights into the neural networks that respond to a specific physiological signal that inhibits subsequent food intake in humans. Previous neuroimaging studies have identified some of the regions found in our whole-brain analysis (OFC, anterior cingulate, insula, putamen, prefrontal areas, the left inferior parietal lobule, right cerebellar hemisphere) in food-related paradigms including examination of the fasted state, glucose ingestion, the fasted to fed transition and in conditions of satiety<sup>6–10,24–27</sup>. However, in general these studies have attributed the observed neural changes to the sensory, reward and cognitive responses to eating or other food-relevant stimuli<sup>22–24,28</sup>. In contrast, we show that administration of a single gut hormone modulates the activity of the OFC and other corticolimbic and higher cortical regions, independently of meal-related sensory experiences. Furthermore, on the PYY day we find that change in neural activity in the left OFC predicts subsequent consummatory behaviour, suggesting that rewarding aspects of food are diminished by PYY modulation of the OFC. In support of our



**Figure 3 | Hypothalamic and OFC response to PYY and predictive effect on subsequent caloric intake.** **a, b,** Group mean time series of change in signal (%) extracted from posterior hypothalamic region (**a**) and left caudolateral OFC (**b**) on saline (green, open, inverted triangles) and PYY (black, open circles) study days and plasma PYY concentrations on saline (blue, filled circles) and PYY (red, filled squares) study days. Data are mean  $\pm$  s.e.m. ( $n = 8$  per group). **c–f,** Correlation plots of signal change (%) from the peak voxel in the posterior hypothalamus (**c** and **e**) and left caudolateral OFC (**d** and **f**) and caloric intake for each subject on saline (**c, d**, inverted, open, green triangles) and PYY study days (**e, f**, open, black circles). Lines indicate linear regression. Corresponding  $r$  and  $P$  values are displayed on the plots.

hypothesis, we find that change in OFC activation negatively correlated with meal pleasantness, a hedonic rating, on the PYY day ( $r = -0.77$ ,  $P = 0.023$ , Supplementary Fig. 11) but not the saline day ( $r = -0.1$ ,  $P = 0.80$ , Supplementary Fig. 11). Together, these findings support the suggestion that the bottom-up processing of internal state (homeostasis) initiated in subcortical structures interfaces with the cognitive level within the OFC<sup>29</sup>, and provide further insight into the mechanisms by which degenerative lesions of the OFC lead to pathological eating behaviour<sup>30</sup>.

Our observations are particularly important as it is clear from the current obesity epidemic that homeostatic processes regulating food intake in humans can be easily over-ridden. Therefore, a key strategy to treat obesity will be to identify agents that modulate the hedonic aspects of food intake. Our studies demonstrate the ability of fMRI-based physiological approaches to provide objective readings of the circuits that regulate feeding that may not be accessible to simple subjective measures. These techniques may be useful in determining whether pharmacological or other interventions have an impact on appropriate neuronal substrates and could be used to predict the efficacy of such treatments. Our findings, coupled with advances gained from rodent studies, will start to bridge the gap between the molecular basis of the neural pathways regulating energy homeostasis and the behavioural influences on human obesity, and may lead to a new understanding of the pathogenesis and treatment of this condition.

## METHODS SUMMARY

BOLD sensitive magnetic resonance images (3.0-Tesla, General Electric Medical Systems Excite II scanner) were collected from eight, healthy, male, right-handed volunteers in a counterbalanced, double-blind, placebo-controlled, crossover study in which intravenous saline or physiologically relevant concentrations of PYY were administered (Supplementary Fig. 1). Scanning started simultaneously with the infusion of saline (from  $t_{-10}$  to  $t_0$ ). At  $t_0$  the infusion was changed to either saline or PYY for a further 90 min and scanning was performed throughout this period ( $t_0$  to  $t_{90}$ ). Throughout the scanning period subjects responded to VAS questions to assess food- and non-food-related feelings. Thirty minutes after the end of the infusion subjects were offered a large meal such that all appetites could be satisfied and caloric intake calculated. Blood samples were taken at 10-min intervals throughout the scanning period ( $t_{-10}$  to  $t_{90}$ ) and then before the meal ( $t_{120}$ ). Images were processed using Statistical Parametric Mapping Software version 5 (SPM5; <http://www.fil.ion.ucl.ac.uk/spm/>). Brain regions that co-varied positively and negatively with plasma PYY concentrations were identified using separate, subject-specific models containing: (1) a regressor for plasma PYY concentration; (2) scan-to-scan subject movement parameters (from the affine transformation); (3) the mean global signal at each time point. Group PYY covariance was examined in a repeated measures ANOVA model. Region of interest (ROI) time series were extracted from local maxima within 4 mm of group activation map X, Y and Z coordinates, and signal change examined for correlation with caloric intake on the PYY and saline study days in the hypothalamus and caudolateral OFC. PYY was quantified by radioimmunoassay and ELISA kits were used to quantify active ghrelin, leptin and insulin, as previously described<sup>18</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 31 May; accepted 31 August 2007.**

**Published online 14 October 2007.**

- Morton, G. J., Cummings, D. E., Baskin, D. G., Barsh, G. S. & Schwartz, M. W. Central nervous system control of food intake and body weight. *Nature* **443**, 289–295 (2006).
- Schwartz, M. W. & Porte, D. Jr. Diabetes, obesity, and the brain. *Science* **307**, 375–379 (2005).
- Berthoud, H. R. Neural control of appetite: cross-talk between homeostatic and non-homeostatic systems. *Appetite* **43**, 315–317 (2004).
- Fulton, S., Woodside, B. & Shizgal, P. Modulation of brain reward circuitry by leptin. *Science* **287**, 125–128 (2000).
- Saper, C. B., Chou, T. C. & Elmquist, J. K. The need to feed: homeostatic and hedonic control of eating. *Neuron* **36**, 199–211 (2002).
- Liu, Y., Gao, J. H., Liu, H. L. & Fox, P. T. The temporal response of the brain after eating revealed by functional MRI. *Nature* **405**, 1058–1062 (2000).
- Small, D. M., Zatorre, R. J., Dagher, A., Evans, A. C. & Jones-Gotman, M. Changes in brain activity related to eating chocolate: from pleasure to aversion. *Brain* **124**, 1720–1733 (2001).
- Tataranni, P. A. et al. Neuroanatomical correlates of hunger and satiation in humans using positron emission tomography. *Proc. Natl Acad. Sci. USA* **96**, 4569–4574 (1999).
- Smeets, P. A. et al. Effect of satiety on brain activation during chocolate tasting in men and women. *Am. J. Clin. Nutr.* **83**, 1297–1305 (2006).
- DelParigi, A. et al. Persistence of abnormal neural responses to a meal in postobese individuals. *Int. J. Obes. Relat. Metab. Disord.* **28**, 370–377 (2004).
- Cummings, D. E. & Overduin, J. Gastrointestinal regulation of food intake. *J. Clin. Invest.* **117**, 13–23 (2007).
- Badman, M. K. & Flier, J. S. The gut and energy balance: visceral allies in the obesity wars. *Science* **307**, 1909–1914 (2003).
- Batterham, R. L. et al. Gut hormone PYY<sub>3–36</sub> physiologically inhibits food intake. *Nature* **418**, 650–654 (2002).
- Batterham, R. L. et al. Inhibition of food intake in obese subjects by peptide YY<sub>3–36</sub>. *N. Engl. J. Med.* **349**, 941–948 (2003).
- Wren, A. M. et al. Ghrelin enhances appetite and increases food intake in humans. *J. Clin. Endocrinol. Metab.* **86**, 5992–5995 (2001).
- Chan, J. L., Stoyneva, V., Kelesidis, T., Raciti, P. & Mantzoros, C. S. Peptide YY levels are decreased by fasting and elevated following caloric intake but are not regulated by leptin. *Diabetologia* **49**, 169–173 (2006).
- le Roux, C. W. et al. Attenuated peptide YY release in obese subjects is associated with reduced satiety. *Endocrinology* **147**, 3–8 (2006).
- Batterham, R. L. et al. Critical role for peptide YY in protein-mediated satiation and body-weight regulation. *Cell Metab.* **4**, 223–233 (2006).
- Abbott, C. R. et al. The inhibitory effects of peripheral administration of peptide YY<sub>3–36</sub> and glucagon-like peptide-1 on food intake are attenuated by ablation of the vagal-brainstem-hypothalamic pathway. *Brain Res.* **1044**, 127–131 (2005).
- Abizaid, A. et al. Ghrelin modulates the activity and synaptic input organization of midbrain dopamine neurons while promoting appetite. *J. Clin. Invest.* **116**, 3229–3239 (2006).
- Fulton, S. et al. Leptin regulation of the mesoaccumbens dopamine pathway. *Neuron* **51**, 811–822 (2006).
- Kringelbach, M. L. Food for thought: hedonic experience beyond homeostasis in the human brain. *Neuroscience* **126**, 807–819 (2004).
- Kringelbach, M. L. The human orbitofrontal cortex: linking reward to hedonic experience. *Nature Rev. Neurosci.* **6**, 691–702 (2005).
- Kringelbach, M. L., O'Doherty, J., Rolls, E. T. & Andrews, C. Activation of the human orbitofrontal cortex to a liquid food stimulus is correlated with its subjective pleasantness. *Cereb. Cortex* **13**, 1064–1071 (2003).
- Wang, G. J. et al. Exposure to appetitive food stimuli markedly activates the human brain. *Neuroimage* **21**, 1790–1797 (2004).
- Gautier, J. F. et al. Effect of satiation on brain activity in obese and lean women. *Obes. Res.* **9**, 676–684 (2001).
- Hinton, E. C. et al. Neural contributions to the motivational control of appetite in humans. *Eur. J. Neurosci.* **20**, 1411–1418 (2004).
- Rolls, E. T. The functions of the orbitofrontal cortex. *Brain Cogn.* **55**, 11–29 (2004).
- DelParigi, A. et al. Successful dieters have increased neural activity in cortical areas involved in the control of behavior. *Int. J. Obes. (Lond.)* **31**, 440–448 (2006).
- Whitwell, J. L. et al. VBM signatures of abnormal eating behaviours in frontotemporal lobar degeneration. *Neuroimage* **35**, 207–213 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank the subjects who participated in this study. We also thank J. Jones and E. Robins for assistance with RIAs; Y. Shaham and H.-R. Berthoud for discussions; and C. Andrew, D. Gasston and the radiography team for technical assistance. This work was supported by the Medical Research Council, Rosetrees Trust and the Travers' Legacy. R.L.B. is an MRC Clinician Scientist.

**Author Contributions** All authors designed the research. R.L.B., J.M.R. and F.O.Z. undertook the patient studies. R.L.B. analysed the plasma samples. F.O.Z., J.M.R., R.L.B. and D.H.F. analysed the fMRI data. All authors contributed to writing the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to R.L.B. ([r.batterham@ucl.ac.uk](mailto:r.batterham@ucl.ac.uk)).

## METHODS

**Subjects.** Eight, healthy, male, right-handed volunteers were recruited for this study. Subjects had a mean age of  $29.6 \pm 2.1$  years, were of normal weight (body mass index =  $21.7 \pm 0.7 \text{ kg m}^{-2}$ ) and were weight stable for the 3 months before recruitment. Smokers, subjects taking any medication, those with psychiatric or neurological illnesses, claustrophobia, metal implants or food allergies were excluded. All volunteers provided informed consent and all aspects of the study were approved by the Research Ethics Committees at University College London and Institute of Psychiatry, London. Before the first scanning session, subjects took part in a simulated scan in a mock MRI scanner to familiarize themselves with the scanning environment. During this visit subjects were also familiarized with the ten visual analogue scales (VAS) used for assessing subjective feelings during the study. Four of the VAS assessed food-related feelings<sup>13–15,18,31,32</sup>: How hungry do you feel right now? How sick do you feel right now? How much food do you think you could eat right now? How pleasant would it be to eat right now? The remaining six VAS assessed non-food-related feelings of internal state<sup>33,34</sup>: How anxious do you feel right now? How warm do you feel right now? How dizzy do you feel right now? How weak do you feel right now? How irritable do you feel right now? How sleepy do you feel right now? Each VAS was 100 mm in length with the text expressing the most positive and the negative rating anchored at each end.

**Study protocol.** The study was performed in a counterbalanced, double-blind, placebo-controlled, crossover manner with each subject studied on two separate occasions 7 days apart (Supplementary Fig. 1). Half of the subjects received PYY on their first study day, whereas the other half received placebo (saline) on their first study day. On the day before each study day, subjects consumed a similar meal between 19:00 and 20:00 and then fasted and drank only water. For the 24-h period preceding each experimental session the participants refrained from alcohol and from strenuous exercise. On each study day subjects arrived at the scanning centre at 9:00, at which time cannulae were inserted into veins in both forearms: right forearm for collection of blood and left forearm for the infusion of PYY or saline. Subjects then relaxed for 60 min before being transferred to the scanner room.

Structural scans were performed before the main BOLD scan. Image acquisition for the BOLD scan started simultaneously with the infusion of saline (from  $t_{-10}$  to  $t_0$ ). At  $t_0$  the infusion was changed to either saline or PYY ( $0.8 \text{ pmol kg}^{-1} \text{ min}^{-1}$ ) for a further 90 min ( $t_0$  to  $t_{90}$ ). Continuous scanning took place from  $t_{-10}$  to  $t_{90}$ . Throughout the scanning period subjects provided VAS measures responding to questions as detailed above. VAS interrogations were back-projected from an LCD projector on to a screen and viewed within the scanner using a mirror system. Subjects rated their feelings on a scale of 0 (low) to 100 (high), marked at every tenth division, by moving a joystick placed in their right hand. The VAS appeared at 1-min intervals with each question repeated every 10 min. On each occasion the cursor appeared in a random initial position along the VAS and the subject had 8 s to respond by moving it to the position representing their rating. The response time for each of the VAS was synchronized with the idle time of the sparse image data acquisition (see below) so as to minimize any possible effect of movement on the images. At  $t_{90}$  (end of the infusion) subjects left the scanner room and were allowed to relax. Thirty minutes after the end of the infusion (that is, at  $t_{120}$ ) the subjects were offered a large buffet lunch such that all appetites could be satisfied. Food and water were quantified pre- and post-prandially and caloric intake calculated. Blood samples were taken at 10-min intervals throughout the scanning period ( $t_{-10}$  to  $t_{90}$ ) and then before the meal ( $t_{120}$ ). Subjects were asked to respond to the ten VAS questions at  $t_{120}$ , before their meal, and immediately after completing their meal. In addition, they were asked to rate the pleasantness of the meal that they had just consumed.

**PYY.** PYY (Lot number FYY3360401R1), obtained from Bachem, was manufactured to Good Manufacturing Practice as specified in the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. PYY was aliquoted and endotoxin tested by the Pharmacy Quality Control Laboratory at the Royal Free Hampstead NHS Trust, London, UK. PYY was dissolved in 0.9% saline (Bayer) containing 5% by volume Haemaccel (Beacon) to reduce adsorption to the syringe and tubing.

**Image acquisition.** Images were acquired in a 3.0-Tesla, General Electric Medical Systems Excite II scanner. Three-hundred image volumes (each consisting of 60 near-axial slices) were collected in a sparse manner using an echo planar imaging (EPI) sequence with an acquisition window of 5 s and a gap of 15 s between the end of one acquisition and the start of the next. This yielded an effective repetition time of 20 s. The 15-s gap was necessary to allow real-time image reconstruction of the image data. It also served to reduce the total number of volumes to be analysed and provided an 'idle' time during which the VAS measures could be collected without directly affecting the magnetic resonance

signal. Echo time was 30 ms. Each image slice was acquired using a  $96 \times 96$  image matrix, zero-filled during reconstruction to yield a final matrix size of  $128 \times 128$ , over a 24-cm field of view. The resulting in-plane pixel size of the images was  $1.875 \times 1.875 \text{ mm}$ . The image slices had a thickness of 2.2 mm with no gap between them. We chose the minimum allowed slice thickness to maximize our ability to identify small brain structures and to reduce the signal dropout owing to through-plane de-phasing effects. Parallel imaging (using the manufacturer's 'ASSET' methodology) was used throughout this study because with this option the duration of the acquisition readout was significantly reduced. This feature allowed us to minimize the propagation of the magnetic-susceptibility-induced phase loss during the EPI readout window. As a result, image distortion in medial temporal lobe regions, orbitofrontal areas and brainstem<sup>35</sup> were significantly reduced. The total acquisition time of the time series was 100 min. In addition, a three-dimensional spoiled gradient echo volumetric scan (SPGR) was obtained from each subject before the collection of the functional BOLD series. The SPGR image had an image matrix size of  $256 \times 256 \times 196$  voxels, with isotropic resolution of 1.1 mm (echo time/repetition time/inversion time =  $2.812/6.616/450 \text{ ms}$ , excitation flip angle  $20^\circ$ ).

**Image pre-processing and analysis.** Images were processed using Statistical Parametric Mapping Software version 5 (SPM5, Wellcome Trust Centre for Neuroimaging, Institute of Neurology, UCL, London, UK; <http://www.fil.ion.ucl.ac.uk/spm/>). Each volume in the time series was realigned within the PET module (to avoid spin excitation history correction, which was unnecessary in this case due to the long repetition time), normalized to the canonical Montreal Neurological Institute (MNI) EPI template with a customized bounding box extended to the lower brain stem ( $Z = -70$ ), and spatially smoothed with a gaussian kernel of  $4 \times 4 \times 4 \text{ mm}$ .

**PYY analysis.** Separate, subject-specific models were generated for each scanning session (saline or PYY). Each model contained: (1) a regressor for plasma PYY concentration (a 300 time-point cubic spline derived from the 11 PYY blood samples collected during the scanning period); (2) scan-to-scan subject movement parameters (from the affine transformation); (3) the mean global signal at each time point (Supplementary Fig. 12). This signal was included to take account of confounds due to low frequency signal changes throughout the scan (receiver drift, time-dependent thermal noise, gradient amplifier heating, aliased physiological signals, and so on). The influence of different high-pass filters was examined using a cut-off of 3,000 s (scan period/2) to emphasise faster changes in PYY concentration while removing residual slower drifts not accounted for by the global signal regressor, and a cut-off of 12,000 s (scan period  $\times 2$ ) to leave the PYY regressor (and residual noise) unaltered. Both filters resulted in a similar pattern of results and only the 3,000 s filtered data are presented here. Serial correlations were estimated with an autoregressive AR (1) model. As can be seen in Supplementary Fig. 12, the average temporal course of the global mean signal has a small net percentage change in magnitude and a temporal profile that differs from that of PYY concentration, suggesting that our results cannot be explained by residual low-frequency confounds. The temporal profile of deep white matter was similar to that of the global mean signal. Group covariance maps for PYY were derived from a second-level analysis consisting of a one-way repeated measures ANOVA with a 2-level factor: infusion (saline, PYY). This second-level approach further attenuated residual noise as it required activated brain regions to have a different pattern of response on each day (for example, an increase in BOLD signal on the PYY day and a decrease in BOLD signal on the saline day). Brain areas co-varying with PYY were considered significant at  $P < 0.05$  corrected for multiple comparisons at the cluster level. This approach is based on the probability of finding a given number of activated neighbouring voxels in a gaussian field<sup>36</sup>. Activation coordinates were converted from MNI to Talairach space<sup>37</sup>.

**Ghrelin analysis.** We carried out a separate analysis of the functional data in which the plasma PYY concentration regressor (1) above was replaced by the ghrelin plasma concentration regressor. These two analyses were performed separately as PYY and ghrelin plasma concentrations regressors were negatively correlated (Fig. 1a and Supplementary Fig. 2d) so that the effect of one partially co-varied out the other when both were included in the first-level models. This was the case even though the change in plasma PYY concentrations was larger than the changes observed in plasma ghrelin concentrations, as SPM is indifferent to regressor scale. The ghrelin regressor was a 300-point cubic spline derived from the 11 blood samples collected during the scan.

**Region of interest analysis.** Regions identified a priori from rodent studies (hypothalamus, substantia nigra, nucleus accumbens, solitary nucleus and tract, parabrachial nucleus) were defined on the group average EPI image using the Talairach atlas and the Atlas of the Human Brainstem<sup>38</sup> (Supplementary Fig. 3). Region of interest (ROI) time series were extracted from local maxima within 4 mm of the group activation map X, Y and Z coordinates. These time series are of the percentage difference in BOLD from the global mean with percentage



change in BOLD over the time series = (mean % BOLD  $t_{80} - t_{90}$ ) - (mean % BOLD  $t_{-10} - t_0$ ).

**VAS fMRI analysis.** Subject-specific models were constructed using an event-related design with one regressor for each VAS question type (convolved with the haemodynamic response function), a regressor for global signal change and a regressor for PYY concentration. The high-pass filter was set at 3,000 s for consistency with the PYY analysis. Activations related to food questions (How hungry do you feel right now? How pleasant would it be to eat right now? How much do you think you could eat right now?) were examined using the one-way repeated measures ANOVA approach described above, with a threshold at  $P < 0.05$  cluster-level corrected. Food-specific VAS activations were examined using the same approach for the contrast food greater than non-food VAS questions.

**Biochemical analysis.** Blood was collected into EDTA tubes containing 5,000 kallikrein inhibitor units of aprotinin (Bayer) per ml at 10-min intervals. Plasma was separated immediately by centrifugation at 4 °C. Plasma samples for subsequent PYY analysis had DPPIV inhibitor (Linco Research Inc.) added to give a final concentration of 100 µM. Samples for analysis of active ghrelin had 100 µl of 1 M HCL added per ml of plasma. For all assays samples were measured in duplicate. ELISA kits were used to quantify active ghrelin, leptin and insulin (Linco Research Inc.). PYY was quantified by radio-immunoassay (Linco Research Inc.). Plasma glucose was measured using a glucose oxidase method (Thermo Scientific).

**Statistical analysis.** Data are presented as mean  $\pm$  s.e.m. unless otherwise indicated, and were analysed with SPSS version 14 (SPSS Inc.). Caloric intake and plasma hormone concentrations are expressed as mean  $\pm$  s.e.m. Integrated area under the curve (AUC) was calculated using the trapezoid rule. Caloric intake was compared using a paired  $t$ -test. Plasma concentrations of PYY, active ghrelin, glucose, insulin and leptin were compared using repeated measures ANOVA with post-hoc Newman–Keuls multiple comparison test. Visual analogue scores and AUCs were compared using Wilcoxon signed rank matched-pair tests. Non-parametric behavioural data obtained from VAS were analysed using Wilcoxon signed rank test. Correlation was assessed using Pearson's product-moment correlation coefficient. For all statistical analyses  $P < 0.05$  was considered significant.

31. Flint, A., Raben, A., Blundell, J. E. & Astrup, A. Reproducibility, power and validity of visual analogue scales in assessment of appetite sensations in single test meal studies. *Int. J. Obes. Relat. Metab. Disord.* **24**, 38–48 (2000).
32. Batterham, R. L. *et al.* Pancreatic polypeptide reduces appetite and food intake in humans. *J. Clin. Endocrinol. Metab.* **88**, 3989–3992 (2003).
33. Rosenthal, J. M. *et al.* The effect of acute hypoglycemia on brain function and activation: a functional magnetic resonance imaging study. *Diabetes* **50**, 1618–1626 (2001).
34. Rosenthal, M. J. *et al.* Caffeine restores regional brain activation in acute hypoglycaemia in healthy volunteers. *Diabet. Med.* **24**, 720–727 (2007).
35. Deichmann, R., Gottfried, J. A., Hutton, C. & Turner, R. Optimized EPI for fMRI studies of the orbitofrontal cortex. *Neuroimage* **19**, 430–441 (2003).
36. Friston, K. J., Holmes, A., Poline, J. B., Price, C. J. & Frith, C. D. Detecting activations in PET and fMRI: levels of inference and power. *Neuroimage* **4**, 223–235 (1996).
37. Talairach, J. & Tournoux, P. *Co-planar Stereotaxic Atlas of the Human Brain* (Thieme, Stuttgart, 1988).
38. Paxinos, G. & Huang, X. *Atlas of the Human Brainstem* (Academic Press, San Diego, 1995).

## LETTERS

# Subcapsular sinus macrophages in lymph nodes clear lymph-borne viruses and present them to antiviral B cells

Tobias Junt<sup>1</sup>, E. Ashley Moseman<sup>1</sup>, Matteo Iannacone<sup>1,3</sup>, Steffen Massberg<sup>1</sup>, Philipp A. Lang<sup>4</sup>, Marianne Boes<sup>2</sup>, Katja Fink<sup>5</sup>, Sarah E. Henrickson<sup>1</sup>, Dmitry M. Shayakhmetov<sup>6</sup>, Nelson C. Di Paolo<sup>6</sup>, Nico van Rooijen<sup>7</sup>, Thorsten R. Mempel<sup>1</sup>, Sean P. Whelan<sup>8</sup> & Ulrich H. von Andrian<sup>1</sup>

Lymph nodes prevent the systemic dissemination of pathogens such as viruses that infect peripheral tissues after penetrating the body's surface barriers. They are also the staging ground of adaptive immune responses to pathogen-derived antigens<sup>1,2</sup>. It is unclear how virus particles are cleared from afferent lymph and presented to cognate B cells to induce antibody responses. Here we identify a population of CD11b<sup>+</sup>CD169<sup>+</sup>MHCII<sup>+</sup> macrophages on the floor of the subcapsular sinus (SCS) and in the medulla of lymph nodes that capture viral particles within minutes after subcutaneous injection. Macrophages in the SCS translocated surface-bound viral particles across the SCS floor and presented them to migrating B cells in the underlying follicles. Selective depletion of these macrophages compromised local viral retention, exacerbated viraemia of the host, and impaired local B-cell activation. These findings indicate that CD169<sup>+</sup> macrophages have a dual physiological function. They act as innate 'flypaper' by preventing the systemic spread of lymph-borne pathogens and as critical gatekeepers at the lymph-tissue interface that facilitate the recognition of particulate antigens by B cells and initiate humoral immune responses.

We have investigated how virus particles that enter peripheral tissues are handled within draining lymph nodes. Hind footpads of mice were injected with fluorescently labelled ultraviolet-inactivated vesicular stomatitis virus (VSV), a cytopathic rhabdovirus that is transmissible by insect bites<sup>3</sup> and elicits T-independent neutralizing B-cell responses<sup>4</sup>. Using multiphoton intravital microscopy (MP-IVM) in popliteal lymph nodes<sup>5</sup> draining the injected footpad, we observed that VSV accumulated in discrete patches on the SCS floor within minutes after subcutaneous injection, whereas the parenchyma and roof of the SCS remained free of virus (Fig. 1a and Supplementary Movie 1). The viral deposits became progressively denser, forming conspicuous irregular reticular patterns, which remained fixed in place for hours.

To characterize the tissue origin of the preferred VSV binding sites in lymph nodes, we reconstituted irradiated Act(EGFP) mice with wild-type bone marrow. The resulting C57BL/6→Act(EGFP) chimaeras expressed enhanced green fluorescent protein (EGFP) in non-haematopoietic cells, particularly lymphatic endothelial cells, on the SCS floor and roof. On injection of fluorescent VSV into the footpad of C57BL/6→Act(EGFP) chimaeras, viral particles flooded the SCS (Supplementary Movie 2). After 3 h, unbound luminal VSV had disappeared, but the SCS floor displayed prominent patches of VSV that did not colocalize with EGFP<sup>+</sup> cells, suggesting

that VSV was captured by haematopoietic cells (Fig. 1b and Supplementary Movie 3). To characterize the putative VSV-capturing leukocytes, we performed electron microscopy on popliteal lymph nodes harvested 5 min after injection of VSV (Fig. 1c). Bullet-shaped, electron-dense VSV particles were selectively bound to discrete regions on the surface of scattered large cells residing within the SCS or just below the SCS floor. VSV-binding cells that were located beneath the SCS floor were typically in contact with the lymph compartment through protrusions that extended into the SCS lumen.

Ultrastructural studies of lymph nodes have shown that the SCS contains many macrophages<sup>6,7</sup>, so we examined whether the VSV-retaining cells belonged to this population. Indeed, confocal microscopy of frozen lymph node sections obtained 30 min after injection into the footpad showed that VSV colocalized in the SCS with a macrophage marker, CD169/sialoadhesin (Fig. 1d). Using flow cytometry, we detected CD169 on about 1–2% of mononuclear cells in lymph nodes, which uniformly expressed CD11b and MHC-II together, indicating that the VSV-binding cells are indeed macrophages (Supplementary Fig. 1). Most CD169<sup>+</sup> cells also expressed other macrophage markers, including CD68 and F4/80, but few expressed the granulocyte/monocyte marker Gr-1. CD169<sup>+</sup> cells also expressed CD11c, but at lower levels than CD11c<sup>high</sup> conventional dendritic cells. We conclude that intact virions enter the lymph within minutes after transcutaneous deposition and accumulate rapidly and selectively on macrophages in the medulla and SCS of draining lymph nodes.

To explore mechanisms for virus fixation, live VSV (20 µg containing  $2 \times 10^8$  plaque-forming units) was injected into hind footpads, and viral titres in draining lymph nodes were assessed 2 h later. There was no defect in VSV retention in draining lymph nodes of mice deficient in complement C3 (Fig. 1e). DH-LMP2a mice, which lack secreted immunoglobulins, had decreased virus titres in spleen but not in popliteal lymph nodes (Fig. 1f). VSV fixation in lymph nodes therefore occurs by means of a mechanism distinct from that used by splenic marginal-zone macrophages, which require C3 and natural antibodies to capture blood-borne VSV<sup>8,9</sup>. Conceivably, the VSV surface glycoprotein may be recognized in lymph nodes by macrophage-expressed carbohydrate-binding scavenger receptors<sup>10</sup>, but the precise mechanism will require further investigation.

What are the consequences of viral capture by macrophages for virus dissemination and antiviral immunity? To address this question, we depleted lymph-node-resident macrophages by injection of

<sup>1</sup>Immune Disease Institute and Department of Pathology, Harvard Medical School, <sup>2</sup>Department of Dermatology, Brigham and Women's Hospital, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA. <sup>3</sup>Immunopathogenesis of Liver Infections Unit, San Raffaele Scientific Institute, Via Olgettina 58, 20132 Milan, Italy. <sup>4</sup>Institute of Experimental Immunology, Zurich University Hospital, Schmelzbergstrasse 12, CH-8091 Zurich, Switzerland. <sup>5</sup>Novartis Institute of Tropical Diseases, 10 Biopolis Road, 138670 Singapore. <sup>6</sup>Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, Washington 98195, USA. <sup>7</sup>Vrije Universiteit, VUMC, Department of Molecular Cell Biology, Faculty of Medicine, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands. <sup>8</sup>Department of Microbiology and Molecular Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA.

clodronate liposomes (CLL) into the footpad<sup>11</sup>. At the dose used, subcutaneously injected CLL selectively eliminated macrophages in lymph nodes draining the injection site, including the popliteal, inguinal and para-aortic lymph nodes<sup>11</sup>, whereas macrophages in distal lymph nodes and spleen were spared (Supplementary Fig. 2a, b). Among the different lymph-node-resident CD11b<sup>+</sup>MHCII<sup>+</sup> phagocytes, CLL preferentially removed the CD169<sup>+</sup> subset, whereas LYVE-1<sup>+</sup> cells and conventional dendritic cells remained unchanged. CLL-treated popliteal lymph nodes had increased B-cell numbers and enlarged follicles seven days after treatment, but other morphological parameters, for example demarcation of the T/B border and the SCS ultrastructure, remained unaltered (Supplementary Fig. 2c–e).

In comparison with untreated lymph nodes, we recovered about tenfold lower viral titres from the draining lymph nodes of CLL-treated mice (Fig. 1g), suggesting that macrophage depletion rendered lymph filtration inefficient. Indeed, VSV titres were markedly increased in blood, spleen and non-draining lymph nodes of CLL-treated mice. Viral dissemination from the injection site to the blood depended strictly on lymph drainage, because circulating VSV was undetectable when virus was injected into footpads of mice that carried an occluding catheter in the thoracic duct, even in CLL-treated mice. Viral titres were low but detectable in thoracic-duct

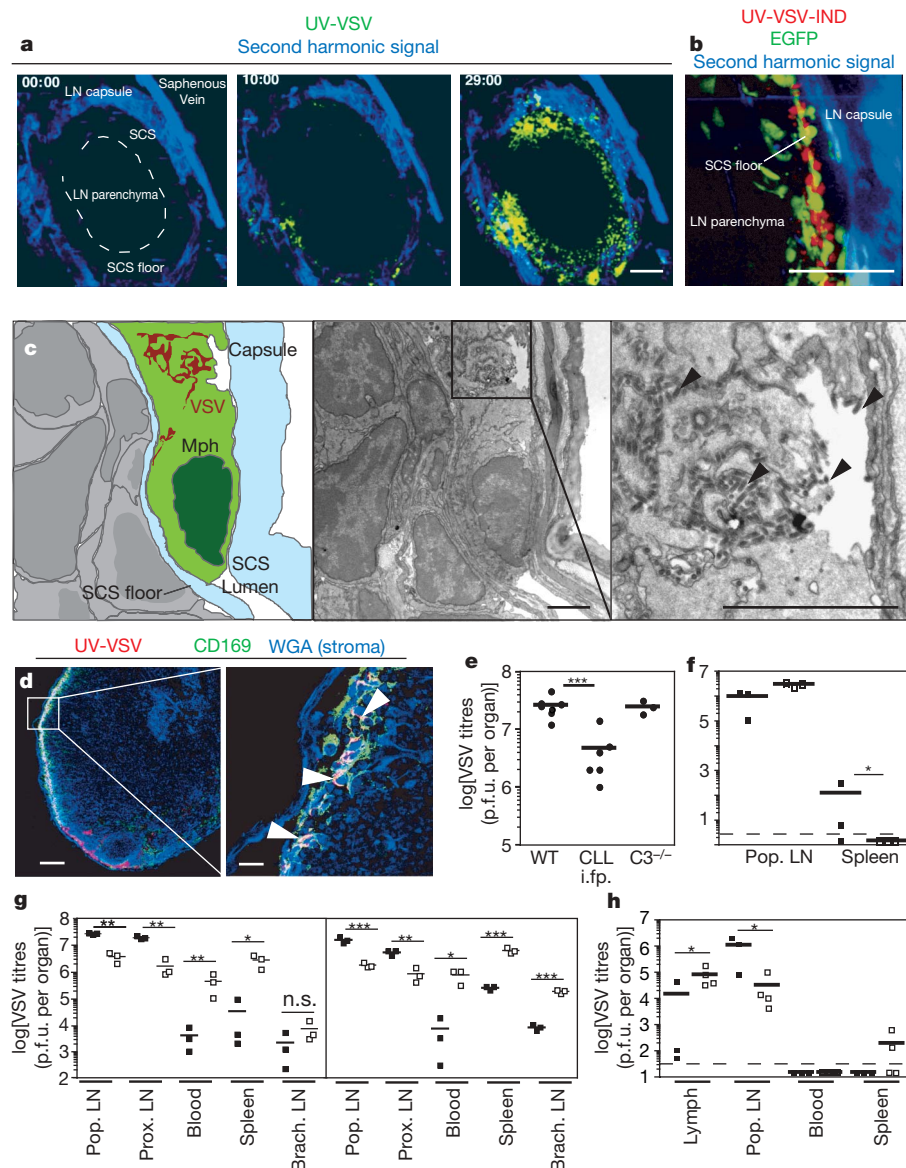
lymph fluid of untreated mice, but increased significantly in CLL-treated animals (Fig. 1h). This indicates that the principal conduit for early viral dissemination from peripheral tissues is the lymph, which is monitored by lymph-node-resident, CLL-sensitive macrophages that prevent the systemic spread of lymph-borne VSV.

This capture mechanism was not unique to VSV: CD169<sup>+</sup> SCS macrophages also retained adenovirus (Supplementary Fig. 3a–c) and vaccinia virus (Supplementary Fig. 3d), indicating that macrophages act as guardians against many structurally distinct pathogens. In contrast, virus-sized latex beads (200 nm diameter) were poorly retained in the SCS after injection into the footpad (Supplementary Fig. 3e). SCS macrophages therefore discriminate between lymph-borne viruses and other particles of similar size. Fluorescent VSV, adenovirus and vaccinia virus also accumulated in the medulla of draining lymph nodes, where they were bound not only by CD169<sup>low</sup> cells (Fig. 1d) but also by CD169<sup>−</sup>LYVE-1<sup>+</sup> lymphatic endothelial cells (Supplementary Fig. 3c, d). This was corroborated in CLL-treated lymph nodes, in which VSV accumulated exclusively on medullary LYVE-1<sup>+</sup> cells (Supplementary Fig. 4).

Next we examined how captured VSV is recognized by B cells. Popliteal lymph nodes contain rare B cells in the SCS lumen (Supplementary Fig. 5a), but we found no evidence for virus-binding lymphocytes within the SCS on electron micrographs (not shown).

**Figure 1 | Capture of lymph-borne VSV by SCS**

**macrophages.** **a**, MP-IVM micrographs of VSV (green) in a popliteal lymph node (LN). Numbers indicate minutes after footpad injection. UV-VSV, ultraviolet-inactivated VSV. Scale bar, 100  $\mu$ m. **b**, VSV (red) accumulation in a C57BL/6  $\rightarrow$ Act(EGFP) recipient 3 h after injection. Scale bar, 50  $\mu$ m. **c**, Electron micrographs of VSV in lymph node 5 min after injection. The middle micrograph is shown schematically on the left and at higher magnification on the right. Arrowheads identify VSV particles. Scale bars, 2  $\mu$ m. Mph, macrophage. **d**, Confocal micrographs of VSV-draining lymph node (30 min). Wheatgerm agglutinin (WGA, blue) was used to stain stromal components. Scale bars, 100  $\mu$ m (left), 15  $\mu$ m (right). **e**, VSV titres in popliteal lymph nodes 2 h after injection into wild-type mice (WT), C3-deficient mice or macrophage-depleted WT mice that received CLL injections into a footpad (CLL i.f.p.). Three asterisks,  $P < 0.001$  (two-way ANOVA, Bonferroni's post-test). p.f.u., plaque-forming units. **f**, VSV capture in Balb/c (filled squares) and DH-LMP2a (open squares) mice. Asterisk,  $P < 0.05$  (unpaired  $t$ -test). **g**, VSV titres after footpad injection in untreated (filled squares) and CLL-treated (open squares) mice (one of two similar experiments;  $n = 3$ ). Left, 2 h; right, 6 h. Pop. LN, popliteal lymph node; Prox. LN, inguinal, para-aortic lymph nodes; Brach. LN, brachial lymph node. Asterisk,  $P < 0.05$ ; two asterisks,  $P < 0.01$ ; three asterisks,  $P < 0.001$ ; n.s., not significant (unpaired  $t$ -test). **h**, Viral titres in lymph, spleen and blood after cannulation of the thoracic duct in untreated (filled squares) and CLL-treated (open squares) mice. Asterisk,  $P < 0.05$  (unpaired  $t$ -test). Horizontal bars in **e–h** indicate means.



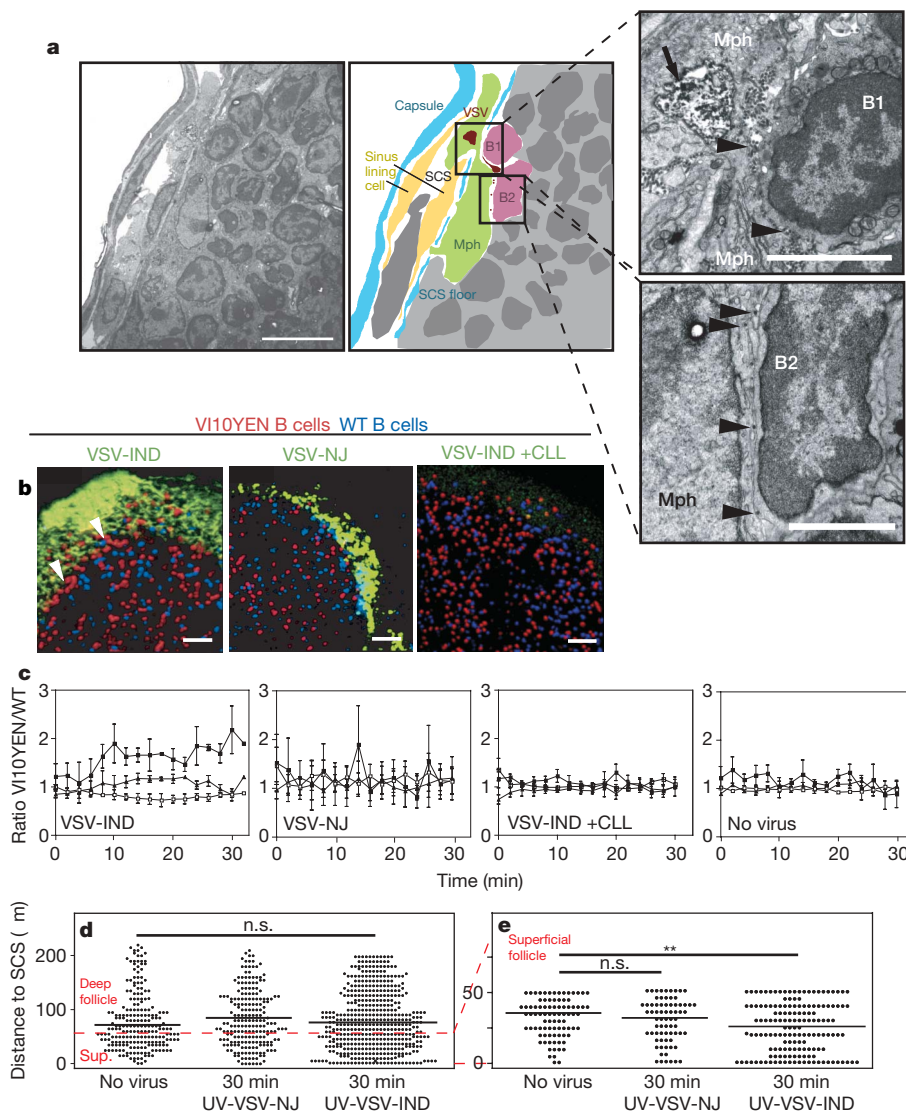


Instead, viral particles were presented to B cells within superficial follicles by macrophages that extended across the SCS floor. After injection of either VSV (Fig. 2a) or adenovirus (Supplementary Fig. 5b–e), virions were readily detectable at B-cell–macrophage interfaces for at least 4 h. This suggested that SCS macrophages shuttle viral particles across the SCS floor for presentation to B cells. Transcytosis seemed unlikely because the few vesicles containing VSV in SCS macrophages showed evidence of viral degradation. In addition, we did not detect substantial motility of virus-binding macrophages by MP-IVM, at least during the first 6 h after challenge. Viral particles therefore most probably reached the lymph node parenchyma by moving along the macrophage surface. VSV and other antigens are also presented to B cells by dendritic cells immigrating from peripheral locations<sup>12,13</sup>, but footpad-derived dendritic cells are not likely to have a function during these very early events because their migration into popliteal lymph nodes takes much longer. We conclude that the SCS floor is not insurmountable for lymph-borne viruses; CD169<sup>+</sup> macrophages seem to act as gatekeepers and facilitators of viral translocation and presentation to B cells.

Next, we explored how naive B cells respond to viral encounter by using two VSV serotypes, Indiana (VSV-IND) and New Jersey (VSV-NJ) (Supplementary Fig. 6)<sup>14</sup>. We compared wild-type B cells with B cells from VI10YEN mice, which express a VSV-IND-specific B cell receptor that does not bind VSV-NJ<sup>15</sup>. By contrast, a small fraction (2–5%) of wild-type B cells bound both serotypes without being

activated. This might reflect low-affinity reactivity with VSV surface glycoprotein or indirect interactions, for example through complement<sup>16</sup>. To assess B cell responses *in vivo*, differentially labelled wild-type and VI10YEN B cells were adoptively transferred and allowed to home to lymph node follicles. Fluorescent ultraviolet-inactivated virus was then injected into footpads, and popliteal lymph nodes were recorded by MP-IVM about 5–35 min later. In virus-free lymph nodes (Supplementary Movie 4) or after injection of VSV-NJ (Supplementary Movie 5), VI10YEN and control B cells had the same distribution (Fig. 2b, c). In contrast, on injection of VSV-IND, VI10YEN cells rapidly accumulated below and within the SCS floor (Supplementary Movie 6). There was no difference in baseline B-cell motility and distribution between CLL-treated and untreated lymph nodes, suggesting that VSV-specific B cells are equally likely to probe the SCS in both conditions (not shown). However, in CLL-treated lymph nodes, fluorescent virus was not retained in the SCS, and VI10YEN B cells failed to congregate in that region, indicating that SCS macrophages are essential for both events (Fig. 2b, Supplementary Movie 7).

To quantify VI10YEN B-cell distribution rigorously, lymph nodes were harvested 30 min after challenge with VSV and analysed by confocal microscopy. Although the entire follicular VI10YEN population retained its overall distribution (Fig. 2d), the subset of cells residing 50  $\mu$ m or less below the SCS shifted towards the SCS in VSV-IND-containing, but not VSV-NJ-containing, lymph nodes (Fig. 2e). It seems unlikely that VI10YEN B cells redistributed to



**Figure 2 | Macrophage-mediated transfer of lymph-borne VSV across the SCS floor alters virus-specific B cell behaviour.** **a**, Electron micrographs and schematic drawing (middle) showing a macrophage (Mph) penetrating the SCS floor of a popliteal lymph node 30 min after injection of VSV. Scale bars, 10  $\mu$ m (left), 4  $\mu$ m (right). The arrow indicates a vacuole with digested VSV; arrowheads indicate virions in the contact zone between the macrophage and B cells. **b**, MP-IVM of polyclonal (blue) and VI10YEN B cells (red) in popliteal lymph nodes. WT, wild type. Scale bars, 50  $\mu$ m. **c**, Regional ratios of VI10YEN B cells/control B cells after VSV injection. Results are from three movies per group. Filled squares, subcapsular sinus; filled triangles, superficial follicle; open squares, deep follicle. Error bars indicate s.e.m. **d, e**, Localization of VI10YEN B cells in popliteal lymph nodes relative to the SCS in the entire follicle (**d**) and in the superficial follicle (**e**). Two asterisks,  $P < 0.01$  (one-way ANOVA with Bonferroni's post-test); n.s., not significant; sup., superficial follicle; UV-VSV, ultraviolet-inactivated VSV. Horizontal bars in **d** and **e** indicate medians.

the SCS because of chemoattractant signals, because unresponsive polyclonal B cells express the same chemoattractant receptors. It is more likely that the random contacts of motile VI10YEN cells with macrophage-bound VSV-IND triggered a B-cell receptor (BCR)-dependent 'stop signal'<sup>17</sup>: short-term exposure to VSV-IND activates LFA-1 and/or  $\alpha 4$  integrins<sup>18</sup> on VI10YEN B cells, resulting in adhesion to the respective ligands, intercellular cell-adhesion molecule-1 (ICAM-1) and vascular cell adhesion molecule-1 (VCAM-1), which are both expressed in the SCS (Supplementary Fig. 7). In addition, VSV-IND bound to SCS macrophages may provide a substrate for VI10YEN B-cell adhesion directly by means of the BCR.

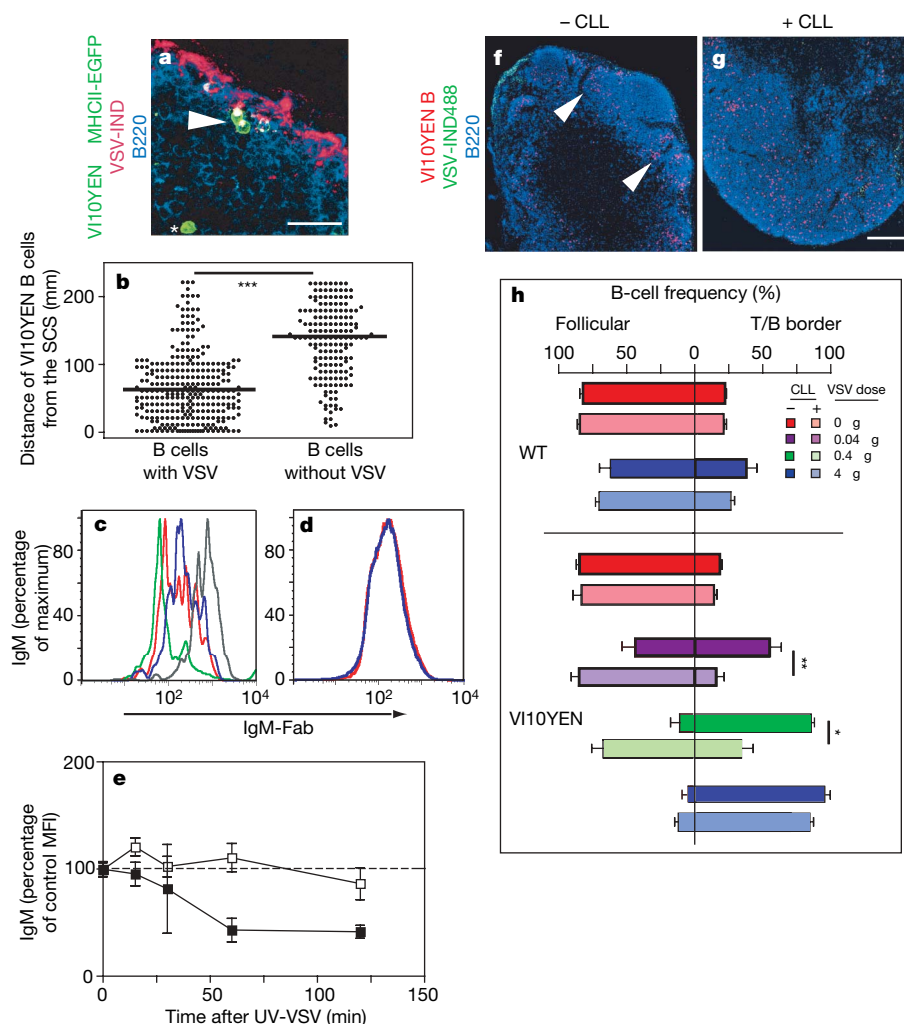
To investigate how captured virions are processed on detection by B cells, we tested B cells from VI10YEN  $\times$  MHCII-EGFP mice, which allowed us to detect endocytosed VSV colocalizing with endosomal MHC-II as an indicator of B-cell priming<sup>19</sup>. Within 30 min after injection, VI10YEN  $\times$  MHCII-EGFP B cells in the superficial follicle had extensively internalized VSV-IND but not VSV-NJ particles (Fig. 3a, b, Supplementary Movie 8, and data not shown). Virus-carrying VSV-specific B cells were infrequent but detectable in deep follicles. These cells may have acquired virions from rare polyclonal B cells that carried VSV on their surface (not shown), or they may correspond to VI10YEN cells that failed to arrest at the SCS after acquiring VSV-IND.

Although our histological findings show that intact virions are preferentially detected and acquired by B cells in the SCS and superficial follicle, MP-IVM measurements of B-cell motility revealed broader antigen dissemination. After injection of VSV-IND, VI10YEN cells responded with a rapid decrease in velocity throughout the entire B follicle (Supplementary Fig. 8). This was observed

equally in CLL-treated and control lymph nodes, indicating that viral antigen reached B cells independently of macrophages. This antigenic material was most probably composed of free viral protein, an inevitable by-product of natural infections. Indeed, purified supernatant of our VSV stocks induced a potent  $\text{Ca}^{2+}$  flux in VI10YEN B cells (Supplementary Fig. 6e). Small lymph-borne proteins are known to diffuse rapidly into follicles and activate cognate B cells<sup>20</sup>. Accordingly, injection of viral supernatant suppressed the motility of follicular VI10YEN B cells without inducing their accumulation at the SCS (not shown), indicating that free VSV surface glycoprotein was contained and active within the viral inoculum. This can explain the macrophage-independent pan-follicular effect of VSV-IND injection.

To determine the kinetics of VI10YEN B cell activation on viral encounter, we measured common activation markers (Supplementary Fig. 9). The co-stimulatory molecule CD86 was first upregulated 6 h after VSV-IND challenge. CD69 was induced more rapidly, but also on polyclonal B cells, presumably by pleiotropic interferon- $\alpha$  signalling<sup>21,22</sup>. Surface IgM (Fig. 3c, d) was downregulated as early as 30 min after challenge, reaching minimum levels within 2 h when more than 70% of VI10YEN cells were BCR<sup>low/neg</sup>. BCR internalization therefore provided the earliest specific readout for virus-specific B-cell activation. VI10YEN B cells in CLL-treated lymph nodes failed to downregulate their BCR during the first 2 h after subcutaneous injection of 20  $\mu\text{g}$  of VSV-IND (Fig. 3e), indicating that SCS macrophages are necessary for efficient early presentation of captured virions to B cells.

Primed B cells eventually solicit help from  $\text{CD4}^+$  T cells<sup>19</sup> for class switch recombination and germinal centre formation. To contact T



**Figure 3 | SCS macrophages are required for early activation of VSV-specific B cells in lymph nodes.** **a**, Confocal micrograph showing MHC-II (green) colocalization with VSV-IND (red; 30 min after injection) in VI10YEN  $\times$  MHCII-EGFP B cells at the SCS (arrowhead) but not the deep follicle (asterisk). Scale bar, 25  $\mu\text{m}$ . **b**, Distance of VSV-associated and VSV-free VI10YEN  $\times$  MHCII-EGFP B cells to the SCS. Horizontal lines indicate medians; three asterisks,  $P < 0.001$ . **c, d**, BCR expression kinetics on VI10YEN cells (**c**) and polyclonal B cells (**d**) after injection of VSV-IND into the footpad. In **c**: dark grey, no virus; blue, 30 min; red, 1 h; green, 2 h. In **d**: red, no virus; blue, 2 h after treatment with ultraviolet-inactivated VSV. **e**, Expression of BCR on VI10YEN cells in CLL-treated (open squares) and untreated (filled squares) popliteal lymph nodes after injection of VSV-IND (20  $\mu\text{g}$ ). Mean fluorescence intensities were normalized to virus-free values (dashed line). UV-VSV, ultraviolet-inactivated VSV. Results are means  $\pm$  s.e.m. (three to five mice). **f, g**, Confocal micrograph of VI10YEN B cells in control (**f**) and CLL-treated popliteal (**g**) lymph nodes 6 h after injection of VSV-IND (0.4  $\mu\text{g}$ ). Scale bar, 125  $\mu\text{m}$ . **h**, Wild-type (WT) and VI10YEN B-cell frequency at T/B borders and in follicles of intact and CLL-depleted lymph nodes, 6 h after injection of VSV-IND. Bars depict the relative frequencies of transferred WT B cells (upper four bars) and VI10YEN cells (lower eight bars) in the two follicular compartments. Colour coding indicates injected doses of ultraviolet-inactivated VSV into CLL-treated recipients (light colours) or untreated recipients (dark colours). Error bars indicate s.e.m.  $n = 3-4$  follicles per two mice; asterisk,  $P < 0.05$ ; two asterisks,  $P < 0.01$ ; three asterisks,  $P < 0.001$  ( $t$ -test).

cells, newly activated B cells migrate towards the T/B border<sup>17,23</sup>. This mechanism operated efficiently in macrophage-sufficient mice: most VI10YEN B cells redistributed to the T/B border within 6 h after injection of as little as 40 ng of VSV-IND into the footpad (Fig. 3f, h, and Supplementary Fig. 10). By contrast, a 100-fold higher viral dose was needed to elicit full redistribution of VI10YEN B cells in CLL-treated mice (Fig. 3g, h). By 12 h after injection, most VSV-specific cells reached the T/B border, irrespective of the injected dose. Thus, even without SCS macrophages, follicular B cells are eventually activated by VSV-derived antigen, although less efficiently.

Thus, we have shown a dual role for CD169<sup>+</sup> macrophages in lymph nodes: they capture lymph-borne viruses preventing their systemic dissemination and they guide captured virions across the SCS floor for the efficient presentation and activation of follicular B cells.

## METHODS SUMMARY

VSV-IND and VSV-NJ virions were purified from culture supernatants of infected BSRT7 cells and used either unmodified or fluorescently labelled with Alexa568 (red) or Alexa488 (green). Fluorescent viruses used for tissue imaging were irradiated with ultraviolet light to prevent the generation of non-fluorescent progeny. Fluorescent labelling or ultraviolet irradiation of VSV-IND particles did not affect their antigenicity or their ability to elicit a Ca<sup>2+</sup> flux in VI10YEN cells (not shown). After injection of fluorescent virus into footpads, draining popliteal lymph nodes were harvested for analysis by electron microscopy or to generate frozen sections for immunostaining and confocal microscopy. To image adoptively transferred B cells in lymph nodes, VI10YEN and wild-type B cells were fluorescently labelled and transferred together by intravenous injection into wild-type or mutant recipient mice. After 18 h, when B cells had homed to B cell follicles, mice were injected with labelled or unlabelled VSV in the right footpad. At different time intervals thereafter, the draining popliteal lymph node was observed by MP-IVM or harvested for confocal microscopy or for flow cytometry to analyse the activation state of virus-specific and control B cells. In some experiments, macrophages in the popliteal lymph node were depleted by subcutaneous injections of CLL, and animals were used for experiments seven to ten days later. MP-IVM, electron microscopy, immunohistochemistry and flow cytometry for various markers were performed on lymph nodes with and without previous CLL treatment. Propagation of VSV from the footpad injection site to the blood and other organs was assessed by injecting a defined amount of live VSV into footpads followed by tissue harvest at 2 or 6 h after VSV injection. To measure viral titres, tissues were homogenized and used in plaque assays. Some viral propagation experiments were performed after cannulation of the thoracic duct.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 29 July; accepted 21 September 2007.

Published online 14 October 2007.

1. von Andrian, U. H. & Mempel, T. R. Homing and cellular traffic in lymph nodes. *Nature Rev. Immunol.* **3**, 867–878 (2003).
2. Karrer, U. *et al.* On the key role of secondary lymphoid organs in antiviral immune responses studied in alymphoplastic (aly/aly) and spleenless (Hox11<sup>-/-</sup>) mutant mice. *J. Exp. Med.* **185**, 2157–2170 (1997).
3. Mead, D. G., Ramberg, F. B. & Mare, C. J. Laboratory vector competence of black flies (Diptera: Simuliidae) for the Indiana serotype of vesicular stomatitis virus. *Ann. NY Acad. Sci.* **916**, 437–443 (2000).
4. Bachmann, M. F., Hengartner, H. & Zinkernagel, R. M. T helper cell-independent neutralizing B cell response against vesicular stomatitis virus: role of antigen patterns in B cell induction? *Eur. J. Immunol.* **25**, 3445–3451 (1995).

5. Mempel, T. R., Henrickson, S. E. & von Andrian, U. H. T-cell priming by dendritic cells in lymph nodes occurs in three distinct phases. *Nature* **427**, 154–159 (2004).
6. Clark, S. L. Jr. The reticulum of lymph nodes in mice studied with the electron microscope. *Am. J. Anat.* **110**, 217–258 (1962).
7. Farr, A. G., Cho, Y. & De Bruyn, P. P. The structure of the sinus wall of the lymph node relative to its endocytic properties and transmural cell passage. *Am. J. Anat.* **157**, 265–284 (1980).
8. Ochsenbein, A. F. *et al.* Protective T cell-independent antiviral antibody responses are dependent on complement. *J. Exp. Med.* **190**, 1165–1174 (1999).
9. Ochsenbein, A. F. *et al.* Control of early viral and bacterial distribution and disease by natural antibodies. *Science* **286**, 2156–2159 (1999).
10. Taylor, P. R. *et al.* Macrophage receptors and immune recognition. *Annu. Rev. Immunol.* **23**, 901–944 (2005).
11. Delemarre, F. G., Kors, N., Kraal, G. & van Rooijen, N. Repopulation of macrophages in popliteal lymph nodes of mice after liposome-mediated depletion. *J. Leukoc. Biol.* **47**, 251–257 (1990).
12. Ludewig, B. *et al.* Induction of optimal anti-viral neutralizing B cell responses by dendritic cells requires transport and release of virus particles in secondary lymphoid organs. *Eur. J. Immunol.* **30**, 185–196 (2000).
13. Qi, H., Egen, J. G., Huang, A. Y. & Germain, R. N. Extrafollicular activation of lymph node B cells by antigen-bearing dendritic cells. *Science* **312**, 1672–1676 (2006).
14. Roost, H. P., Haag, A., Burkhart, C., Zinkernagel, R. M. & Hengartner, H. Mapping of the dominant neutralizing antigenic site of a virus using infected cells. *J. Immunol. Methods* **189**, 233–242 (1996).
15. Hangartner, L. *et al.* Antiviral immune responses in gene-targeted mice expressing the immunoglobulin heavy chain of virus-neutralizing antibodies. *Proc. Natl Acad. Sci. USA* **100**, 12883–12888 (2003).
16. Rossbacher, J. & Shlomchik, M. J. The B cell receptor itself can activate complement to provide the complement receptor 1/2 ligand required to enhance B cell immune responses *in vivo*. *J. Exp. Med.* **198**, 591–602 (2003).
17. Okada, T. *et al.* Antigen-engaged B cells undergo chemotaxis toward the T zone and form motile conjugates with helper T cells. *PLoS Biol.* **3**, e150 (2005).
18. Dang, L. H. & Rock, K. L. Stimulation of B lymphocytes through surface Ig receptors induces LFA-1 and ICAM-1-dependent adhesion. *J. Immunol.* **146**, 3273–3279 (1991).
19. Vascotto, F. *et al.* Antigen presentation by B lymphocytes: how receptor signaling directs membrane trafficking. *Curr. Opin. Immunol.* **19**, 93–98 (2007).
20. Pape, K. A., Catron, D. M., Itano, A. A. & Jenkins, M. K. The humoral immune response is initiated in lymph nodes by B cells that acquire soluble antigen directly in the follicles. *Immunity* **26**, 491–502 (2007).
21. Barchet, W. *et al.* Virus-induced interferon  $\alpha$  production by a dendritic cell subset in the absence of feedback signaling *in vivo*. *J. Exp. Med.* **195**, 507–516 (2002).
22. Shiow, L. R. *et al.* CD69 acts downstream of interferon- $\alpha/\beta$  to inhibit S1P1 and lymphocyte egress from lymphoid organs. *Nature* **440**, 540–544 (2006).
23. Reif, K. *et al.* Balanced responsiveness to chemoattractants from adjacent zones determines B-cell position. *Nature* **416**, 94–99 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank G. Cheng, M. Flynn and D. Baumjohann for technical support; R. M. Zinkernagel and H. Hengartner for providing VI10YEN mice; A. Wagers for providing Act(EGFP) mice; M. Ericsson and E. Benecchi for expert support in electron microscopy studies; S. Behnke for immunohistochemistry; and D. Cureton for help and advice with VSV preparations. This work was supported by grants from the NIH-NIAID (to U.H.v.A.), a Pilot and Feasibility Grant from the Harvard Skin Disease Research Center (to T.J. and U.H.v.A.), a stipend from the Swiss National Science Foundation (to T.J.) and a NIH T32 Training Grant in Hematology (to E.A.M.).

**Author Contributions** T.J. and U.H.v.A. designed the study; T.J., E.A.M., M.I., S.M. and P.A.L. performed experiments; T.J., E.A.M., M.I. and S.M. collected and analysed data; M.B., K.F., N.C.D.P., D.M.S., N.v.R. and S.P.W. provided reagents and mice; T.J., E.A.M., M.I. and U.H.v.A. wrote the manuscript; S.M., K.F., S.E.H., T.M. and S.P.W. gave technical support and conceptual advice.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to U.V.A. ([uva@hms.harvard.edu](mailto:uva@hms.harvard.edu)).



## METHODS

**Mice and antibodies.** C57BL/6 and BALB/c mice were purchased from Taconic Farms. VI10YEN<sup>15</sup>, C3<sup>-/-</sup> (ref. 24), MHCII-EGFP<sup>25</sup>, Act(EGFP)<sup>26</sup> and DH-LMP2A mice<sup>27</sup> were bred in barrier animal facilities at Harvard Medical School and the Immune Disease Institute (IDI). Radiation chimaeras were generated by irradiation of Act(EGFP) mice with two doses of 650 rad and reconstitution with C57BL/6 bone marrow, and were allowed to reconstitute for eight weeks before use. In some experiments, SCS macrophages were depleted by footpad injections of 30 µl clodronate liposomes (CLL), seven to ten days before the experiment. Clodronate is a gift from Roche Diagnostics GmbH. Other reagents for preparation of liposomes were phosphatidylcholine (LIPOID E PC; Lipoid GmbH) and cholesterol (Sigma-Aldrich). Mice were housed under specific pathogen-free and antiviral antibody-free conditions in accordance with National Institutes of Health guidelines. All experimental animal procedures were approved by the Institutional Animal Committees of Harvard Medical School and the IDI. Antibodies were purchased from BD Biosciences, except anti-B220-Alexa647 (Invitrogen-Caltag), anti-LYVE-1 (Millipore-Upstate), goat anti-rabbit-APC (Invitrogen), goat anti-GFP-FITC (Rockland), anti-FITC-Alexa488 (Invitrogen) and Fab anti-IgM-FITC (Jackson ImmunoResearch). The following antibodies were purchased from AbD-Serotec: anti-CD68-Alexa647, anti-CD11b-Alexa647, F4/80-Alexa647 and anti-CD169-FITC (3D6). The anti-idiotypic antibody 35.61 for detection of the VI10 BCR in VI10YEN mice<sup>15</sup> was produced from hybridoma supernatants in accordance with standard methods.

**Flow cytometry.** Flow cytometric analysis of blood samples was performed after retro-orbital phlebotomy of mice and lysis of erythrocytes with ACK buffer (0.15 M NH<sub>4</sub>Cl, 1 mM KHCO<sub>3</sub>, 0.1 mM EDTA (disodium salt), pH 7.2). Single-cell suspensions of lymph nodes and spleens for flow cytometry were generated by careful mincing of tissues and subsequent digestion at 37 °C for 40 min in DMEM (Invitrogen-Gibco) in the presence of 250 µg ml<sup>-1</sup> liberase CI (Roche) plus 50 µg ml<sup>-1</sup> DNase I (Roche). After 20 min of digestion, samples were passed vigorously through an 18-gauge needle to ensure complete organ dissociation. All flow cytometric analyses were performed in FACS buffer containing PBS with 2 mM EDTA and 2% FBS (Invitrogen-Gibco) on a FACScalibur (BD Pharmingen), and analysed by FlowJo software (Treestar Inc.). For Ca<sup>2+</sup> flux, cells were labelled for 90 min at 37 °C with 4 µM Fluo-LOJO (Teflabs) in DMEM containing 10% FCS. Cells were spun through FCS and used immediately.

**Viruses and VSV plaque assay.** VSV, serotypes Indiana (VSV-IND; Mudd-Summers-derived clone, *in vitro* rescued<sup>28</sup> and plaque purified) or New Jersey (VSV-NJ; Pringle Isolate, plaque purified) were propagated at a multiplicity of infection (m.o.i.) of 0.01 on BSRT7 cells. Supernatants of infected cells were cleared from cell debris by centrifugation at 2,000g, filtered through 0.45-µm sterile filters and subjected to ultracentrifugation at 40,000g for 90 min. Pellets were resuspended in PBS and purified by ultracentrifugation (157,000g, 60 min) through a cushion of 10% sucrose in NTE buffer (0.5 mM NaCl, 10 mM Tris-HCl pH 7.5, 5 mM EDTA pH 8). After resuspension overnight in PBS, virus protein was quantified by bicinchoninic acid assay (Pierce), and infectivity was quantified by plaque assay. Some batches were labelled with carboxylic acid succinimidyl esters of AlexaFluor-488 or AlexaFluor-568 (Invitrogen-Molecular Probes) at a 10<sup>4</sup>–10<sup>5</sup>-fold molar excess of Alexa dye over virus particles. Unconjugated dye was removed by ultracentrifugation through 10% sucrose in NTE buffer; pellets were resuspended in PBS and stored frozen. Infectivity of VSV preparations was quantified by plaque assay on green monkey kidney cells (Vero). VSV titres from organs of infected mice were determined similarly, after homogenization of the organs with a Potter–Elvehjem homogenizer. When necessary, during viral preparation, the roughly 4-ml supernatants from the 157,000g ultracentrifugation were collected and concentrated with a 10-kDa molecular mass cut-off Amicon Ultra (Millipore). To account for residual infectivity in concentrated supernatants, VSV stocks were diluted to levels of infectivity equal to that of the concentrated supernatants, and the Ca<sup>2+</sup> flux in VI10YEN B cells was compared over further 100-fold dilutions of VSV and supernatant. Ultraviolet-inactivated, AlexaFluor-568-labelled adenovirus 5 was generated in accordance with standard procedures<sup>29</sup>. All work on infectious materials was performed in designated BL2+ workspaces, in accordance with institutional guidelines, and approved by the Harvard Committee on Microbiological Safety.

**VSV neutralization assay.** Serum from immunized mice was prediluted 40-fold in MEM containing 2% FCS. Serial twofold dilutions were mixed with equal volumes of VSV (500 plaque-forming units ml<sup>-1</sup>) and incubated for 90 min at 37 °C in 5% CO<sub>2</sub>. Serum-virus mixture (100 µl) was transferred to Vero cell monolayers in 96-well plates and incubated for 1 h at 37 °C. The monolayers were overlaid with 100 µl of DMEM containing 1% methylcellulose and incubated for 24 h at 37 °C. Subsequently, the overlay was discarded and the monolayer was fixed and stained with 0.5% crystal violet. The highest dilution of serum

that decreased the number of plaques by 50% was taken as the titre. To determine IgG titres, undiluted serum was pretreated with an equal volume of 0.1 mM 2-mercaptoethanol in saline solution.

**Adhesion assays.** Corning 96-well plates were coated overnight with dilutions of recombinant murine VCAM-1-Fc or ICAM-1-Fc (R&D Systems) or with purified VSV-IND in PBS in triplicate. Negative control wells were coated with 4% BSA, positive control wells were coated with 1 mg ml<sup>-1</sup> poly-(L-lysine). Plates were blocked for 1–2 h at 4 °C with HBSS/1% BSA and washed. Naive B cells from VI10YEN or C57BL/6 mice were negatively selected by magnetic cell separation with CD43 magnetic beads (Miltenyi) and added to the plates at 3 × 10<sup>5</sup> per well in HBSS with 1% BSA, 1 mM Ca<sup>2+</sup> and 1 mM Mg<sup>2+</sup> in the presence or absence of ultraviolet-inactivated VSV-IND (m.o.i. 1,000) for 30 min at 37 °C. After gentle washing three times in HBSS with 1% BSA, plates were fixed for 10 min with PBS/10% glutaraldehyde, stained for 45 min with 0.5% crystal violet/20% methanol and washed in water. The dye was eluted by addition of 1% SDS and the absorbance at 570 nm was determined spectrophotometrically (SpectraMax340PC microplate reader and SoftmaxPro 3.1.2 software; Molecular Devices Corporation) after 30 min.

**Confocal microscopy.** For some analyses, both hind footpads of C57BL/6 mice were injected with 20 µg of AlexaFluor-568-labelled or AlexaFluor-488-labelled VSV-IND or VSV-NJ and the draining lymph nodes were harvested after 30 min. For other experiments, mice were transfused with 10<sup>7</sup> negatively selected naive B cells from VI10YEN × MHCII-EGFP mice one day before the experiment. At predetermined time points, popliteal lymph nodes were fixed *in situ* by footpad injections of phosphate-buffered L-lysine with 1% paraformaldehyde/periodate (PLP). After removal of popliteal lymph nodes and incubation for 3–5 h in PLP at 4 °C, popliteal lymph nodes were washed in 0.1 M PBS pH 7.2 and cryoprotected by an ascending series of 10%, 20% and 30% sucrose in PBS. Samples were snap-frozen in TBS tissue-freezing liquid (Triangle Biomedical Sciences) and stored at –80 °C. Sections of 40 µm thickness were mounted on Superfrost Plus slides (Fisherbrand) and stained with fluorescent antibodies in a humidified chamber after Fc receptor blockade with 1 µg ml<sup>-1</sup> antibody 2.4G2 (BD Pharmingen). Samples were mounted in FluorSave reagent solution (EMD-Calbiochem) and stored at 4 °C until analysis. Images were collected with a Bio-Rad confocal microscopy system using an Olympus BX50WI microscope and 10×/0.4 numerical aperture or 60×/1.2 numerical aperture water-immersion objective lenses. Images were analysed with LaserSharp2000 software (Bio-Rad Cell Science) and Photoshop CS (Adobe). Quantification of B cells localized at the T/B border was done by counting cells that were within 50 µm of the T/B border, as denoted by B220 counterstain; any cells localized in more central regions were considered follicular.

**Electron microscopy.** Popliteal lymph nodes were fixed *in situ* by footpad injection of 2% formaldehyde and 2.5% glutaraldehyde in 0.1 M sodium cacodylate buffer pH 7.4. The lymph nodes were excised and immersed in the same buffer overnight at 4 °C, washed in cacodylate buffer and osmicated with 1% osmium tetroxide/1.5% potassium ferrocyanide (in water) for 1 h at 18–22 °C in the dark. After being washed in water, samples were washed three or four times in 0.05 M maleate buffer pH 5.15. Samples were counterstained for 2 h in 1% uranyl acetate in maleate buffer and washed three times in water. Samples were dehydrated by incubation for 15 min in dilutions of ethanol in water (70%, 90% and 100%), incubated in propylene oxide for 1 h and transferred into Epon mixed 1:1 with propylene oxide at room temperature overnight. Samples were moved to an embedding mould filled with freshly mixed Epon, and heated for 24–48 h at 60 °C for polymerization. Samples were analysed on a Tecnai G2 Spirit BioTWIN electron microscope at the Harvard Medical School electron microscope facility.

**Intravital MP-IVM of the popliteal lymph node.** Naive B cells were negatively selected by magnetic isolation with CD43 beads (Miltenyi). VI10YEN B cells were labelled for 20 min at 37 °C with 10 µM 5-(and 6-)-(((4-chloromethyl)benzoyl)amino)tetramethylrhodamine (CMTMR; Invitrogen), C57BL/6 B cells were labelled for 25 min at 37 °C with 10 µM 7-amino-4-chloromethylcoumarin (CMAC; Invitrogen). In some experiments, labels were swapped between wild-type and VI10YEN B cells to exclude unspecific dye effects (data not shown). B cells ((5–6) × 10<sup>6</sup>) from each population were mixed and adoptively transferred by tail-vein injection into C57BL/6 recipient mice one day before analysis. In some experiments, recipient C57BL/6 mice had received an injection of 30 µl of CLL into the hind footpad 7–10 days before the experiment to eliminate SCS macrophages<sup>11</sup>. At 18 h after adoptive B-cell transfer, recipient mice were anaesthetized by intraperitoneal injection of ketamine (50 mg kg<sup>-1</sup>) and xylazine (10 mg kg<sup>-1</sup>). The right popliteal lymph node was prepared microscopically for MP-IVM and positioned on a custom-built microscope stage as described<sup>5</sup>. Care was taken to spare blood vessels and afferent lymph vessels. The exposed lymph node was submerged in normal saline and covered with a glass coverslip. A thermocouple was placed next to the lymph node to monitor local temperature, which was maintained at 36–38 °C. MP-IVM was performed on a Bio-Rad

2100MP system at an excitation wavelength of 800 nm, from a tunable MaiTai Ti:sapphire laser (Spectra-Physics). Fluorescently labelled VSV (20 µg in 20 µl) was injected through a 31-gauge needle into the right hind footpad of recipient mice concomitant to observation. For four-dimensional offline analysis of cell migration, stacks of 11 optical *x-y* sections with 4-µm *z* spacing were acquired every 15 s with electronic zooming to  $\times 1.8$ – $3.0$  through a  $20\times/0.95$  numerical aperture water-immersion objective lens (Olympus). Emitted fluorescence and second-harmonic signals were detected through 400/40 nm, 450/80 nm, 525/50 nm and 630/120 nm bandpass filters with non-descanned detectors to generate three-colour images. Sequences of image stacks were transformed into volume-rendered, four-dimensional time-lapse movies using Volocity software (Improvision). Three-dimensional instantaneous velocities were determined by semi-automated cell tracking with Volocity and computational analysis by Matlab (Mathworks). Accumulation of cells at the SCS was determined by manual movie analysis performed by blinded observers. Every 2 min, the VI10YEN B cells and polyclonal B cells were counted at the SCS, in the superficial follicle (less than 50 µm distance from the SCS) and the deep follicle (more than 50 µm distance from the SCS), and ratios of VI10YEN/polyclonal B cells were expressed for each compartment in the entire 30-min movie.

**Thoracic duct cannulation.** For thoracic duct cannulation, mice received 200 µl of olive oil by gavage 30 min before cannulation to facilitate visualization of the lymph vessels. Animals were then anaesthetized with xylazine (10 mg kg<sup>-1</sup>) and ketamine HCl (50 mg kg<sup>-1</sup>). A polyethylene catheter (PE-10) was inserted into the right jugular vein for continuous infusion (2 ml h<sup>-1</sup>) of Ringer's lactate (Abbott Laboratories) containing 1 U ml<sup>-1</sup> heparin (American Pharmaceutical Partners). With the use of a dissecting microscope, the thoracic duct was exposed through a left subcostal incision. Silastic silicon tubing (0.012 inch internal diameter; Dow Corning) was flushed with heparinized (50 U ml<sup>-1</sup>) PBS (DPBS; Mediatech), inserted into the cisterna chyli through a roughly 0.3-mm incision and fixed with isobutyl cyanoacrylate monomer (Nexaband; Abbott Laboratories). The remaining part of the tubing was drawn to the exterior through the posterior abdominal wall. Subsequently, the abdominal incision was closed with a 6-0 nonabsorbable running suture (Sofsilik; Tyco Healthcare Group). After equilibration of lymph flow for 30 min, animals were injected into the footpad with 10<sup>8</sup> plaque-forming units of VSV-IND and lymph samples were collected on ice for 6 h. Blood and organs were taken after 6 h of collection of thoracic duct lymph and plaque-assayed as described above. Lymph and organs were plaque-assayed as described above. In some experiments the draining popliteal and para-aortic lymph nodes were surgically excised and the surrounding lymph vessels were cauterized to prevent lymph-borne viral access to the blood.

24. Wessels, M. R. *et al.* Studies of group B streptococcal infection in mice deficient in complement component C3 or C4 demonstrate an essential role for complement in both innate and acquired immunity. *Proc. Natl Acad. Sci. USA* **92**, 11490–11494 (1995).
25. Boes, M. *et al.* T-cell engagement of dendritic cells rapidly rearranges MHC class II transport. *Nature* **418**, 983–988 (2002).
26. Wright, D. E. *et al.* Cyclophosphamide/granulocyte colony-stimulating factor causes selective mobilization of bone marrow hematopoietic stem cells into the blood after M phase of the cell cycle. *Blood* **97**, 2278–2285 (2001).
27. Casola, S. *et al.* B cell receptor signal strength determines B cell fate. *Nature Immunol.* **5**, 317–327 (2004).
28. Whelan, S. P., Ball, L. A., Barr, J. N. & Wertz, G. T. Efficient recovery of infectious vesicular stomatitis virus entirely from cDNA clones. *Proc. Natl Acad. Sci. USA* **92**, 8388–8392 (1995).
29. Leopold, P. L. *et al.* Fluorescent virions: dynamic tracking of the pathway of adenoviral gene transfer vectors in living cells. *Hum. Gene Ther.* **9**, 367–378 (1998).

# A translocation signal for delivery of oomycete effector proteins into host plant cells

Stephen C. Whisson<sup>1</sup>, Petra C. Boevink<sup>1</sup>, Lucy Moleleki<sup>1</sup>, Anna O. Avrova<sup>1</sup>, Juan G. Morales<sup>1</sup>, Eleanor M. Gilroy<sup>1</sup>, Miles R. Armstrong<sup>1</sup>, Severine Grouffaud<sup>1,2</sup>, Pieter van West<sup>2</sup>, Sean Chapman<sup>1</sup>, Ingo Hein<sup>1</sup>, Ian K. Toth<sup>1</sup>, Leighton Pritchard<sup>1</sup> & Paul R. J. Birch<sup>1</sup>

Bacterial<sup>1</sup>, oomycete<sup>2</sup> and fungal<sup>3</sup> plant pathogens establish disease by translocation of effector proteins into host cells, where they may directly manipulate host innate immunity. In bacteria, translocation is through the type III secretion system<sup>1</sup>, but analogous processes for effector delivery are uncharacterized in fungi and oomycetes. Here we report functional analyses of two motifs, RXLR and EER, present in translocated oomycete effectors<sup>2</sup>. We use the *Phytophthora infestans* RXLR-EER-containing protein Avr3a as a reporter for translocation because it triggers RXLR-EER-independent hypersensitive cell death following recognition within plant cells that contain the R3a resistance protein<sup>4,5</sup>. We show that Avr3a, with or without RXLR-EER motifs, is secreted from *P. infestans* biotrophic structures called haustoria, demonstrating that these motifs are not required for targeting to haustoria or for secretion. However, following replacement of Avr3a RXLR-EER motifs with alanine residues, singly or in combination, or with residues KMIK-DDK—representing a change that conserves physicochemical properties of the protein—*P. infestans* fails to deliver Avr3a or an Avr3a–GUS fusion protein into plant cells, demonstrating that these motifs are required for translocation. We show that RXLR-EER-encoding genes are transcriptionally upregulated during infection. Bioinformatic analysis identifies 425 potential genes encoding secreted RXLR-EER class proteins in the *P. infestans* genome. Identification of this class of proteins provides unparalleled opportunities to determine how oomycetes manipulate hosts to establish infection.

One of the most significant unanswered questions in plant pathology is how microbial eukaryotic pathogens, such as fungi and oomycetes, manipulate host metabolism and defences to establish disease. Identification of bacterial effector proteins has provided unparalleled insights into the evolution of bacterial pathogenesis and host mimicry employed by bacterial proteins to interfere with host signalling and signal transduction processes.

In contrast, we know little about effector function in fungi and oomycetes, and nothing about how such proteins are delivered into host cells. Recognition by intracellular host resistance (R) proteins, triggering localized cell death called the hypersensitive response, is evidence that fungal and oomycete effectors enter plant cells. Such effectors are termed avirulence (AVR) proteins, and several have been identified in fungi<sup>6–8</sup> and oomycetes<sup>4,9–11</sup>.

The identification of a common motif, RXLR, in oomycete AVR proteins<sup>2,11</sup> sparked excitement and speculation regarding translocation of effectors from these important filamentous, fungus-like pathogens. The motif resembles, in sequence and relative location, a motif (RXLXE/D/Q) required for translocation of *Plasmodium* (malaria parasite) effectors into host erythrocytes<sup>12,13</sup>. Indeed, the

*P. infestans* RXLR motif has been demonstrated to function in the *Plasmodium* pathosystem<sup>14,15</sup>. Three of the four oomycete AVR proteins also possess an EER motif within 25 residues downstream of the RXLR<sup>11</sup>. Two key questions arise regarding the motifs' role: (1) are they required to translocate effector proteins into plant host cells? (2) Are such effector proteins delivered from haustoria, which, following invagination of the host cell membrane, are in intimate contact with host cells? Our findings are summarized schematically in Supplementary Fig. 1.

*Phytophthora infestans* is the best-studied oomycete as a result of its role in precipitating the Irish potato famines in the mid-19th century, and remains the most economically important potato pathogen. Infection involves two phases: a biotrophic phase up to 36 h post inoculation (h.p.i.) in which *P. infestans* forms haustoria and requires living plant tissue, and an ensuing necrotrophic phase in which infected host tissue becomes necrotic. Gene expression of 38 predicted secreted RXLR-EER class effectors from *P. infestans* was quantified using real-time reverse transcribed polymerase chain reaction (real-time RT–PCR) in pre-infection stages, and throughout infection of potato up to 72 h.p.i. The 38 *P. infestans* genes (see Supplementary Information for bioinformatics analyses) were all induced and could be broadly grouped as: predominantly upregulated in pre-infection only; predominantly in pre-infection and biotrophy; in pre-infection and throughout infection; and, in biotrophy only (Supplementary Fig. 2 and Supplementary Table 1). Upregulation of these genes during infection indicates that the RXLR-EER motifs may have high predictive value in identifying potential secreted pathogenicity determinants. Using a hidden Markov model based on the alignment of the 38 induced genes we queried the preliminary assembly of the *P. infestans* genome sequence ([http://www.broad.mit.edu/annotation/genome/phytophthora\\_infestans](http://www.broad.mit.edu/annotation/genome/phytophthora_infestans)), which predicted 284 RXLR-EER class effectors. We also extended existing methodology<sup>14</sup> to predict 310 RXLR-EER class sequences. These combined approaches predicted 425 sequences, representing a proposed upper limit for the number of secreted RXLR-EER class genes in *P. infestans*. One hundred and sixty-nine sequences were identified by both methods, constituting a set of sequences with priority for further investigation (see Supplementary Information for bioinformatics analyses). These approaches also identified 400 and 314 RXLR-EER class effector genes in the recently reported genome sequences<sup>16</sup> of *P. sojae* and *P. ramorum*, respectively.

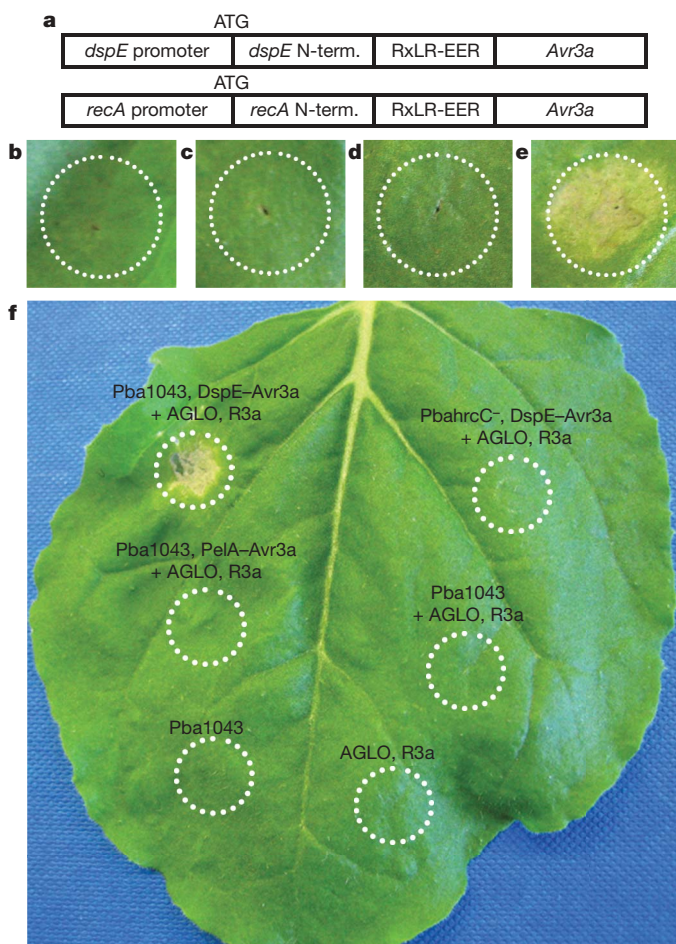
Resistance protein R3a recognizes Avr3a in the host cytoplasm<sup>4</sup>, triggering the hypersensitive response. Intracellular recognition of Avr3a was further investigated here by expression and delivery of Avr3a from the type III secretion system (T3SS) of the bacterial potato pathogen, *Pectobacterium atrosepticum*<sup>17,18</sup>. Secretion of

<sup>1</sup>Plant Pathology Programme, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK. <sup>2</sup>Aberdeen Oomycete Group, University of Aberdeen, Institute of Medical Sciences, Aberdeen AB25 2ZD, UK.



Avr3a as a fusion protein by the T3SS was demonstrated by co-infiltration of *Nicotiana benthamiana* leaves with *Agrobacterium tumefaciens* that delivered the *R3a* gene, and yielded a clear hypersensitive response. In contrast, a hypersensitive response was not observed when Avr3a was fused to a bacterial non-secreted protein or a type II secreted protein (Supplementary Fig. 3), or when the T3SS was disabled, indicating that translocation was dependent on the T3SS (Fig. 1). These results confirm that Avr3a must be delivered inside the plant cell for recognition by R3a, implying that Avr3a translocation by *P. infestans* may require a structure and environment—as is created by the haustorium and extra-haustorial matrix—and/or a mechanism functionally analogous to the T3SS.

The ability of Avr3a to trigger hypersensitive-response-mediated resistance was exploited to determine the role of the RXLR-EER motifs in translocation of the effector into host cells. The RXLR

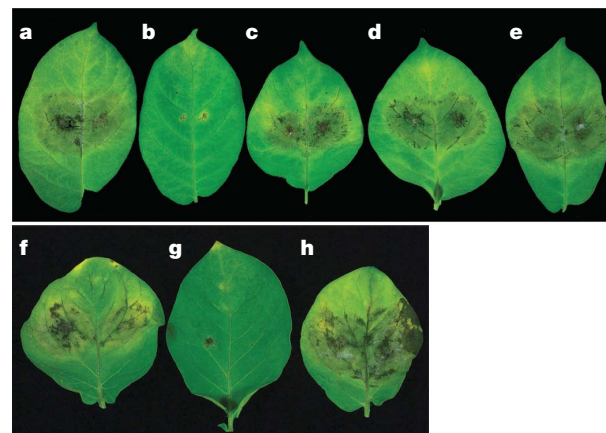


**Figure 1 | Recognition of Avr3a by R3a follows delivery into the host cell by the T3SS of *Pectobacterium atrosepticum* (*Pba*).** **a**, Constructs for expression of Avr3a in *Pba* using promoter regions and N-terminal 100 amino acid (pr100)-coding regions of *dspE* (T3SS-translocated) and *recA* (non-secreted). The N-terminal 50 amino acids of *Pba* DspE (ref. 18) contains the amino acid profile associated with translocation of T3SS effectors in *Pseudomonas syringae*<sup>27</sup>. The translation start site is indicated as ATG. Avr3a fusion constructs encode protein lacking native Avr3a signal peptide (amino acids 1–21), but retaining RXLR-EER motifs. **b**, Infiltration of *A. tumefaciens* (AGLO) delivering R3a. **c**, Infiltration of *Pba* expressing *dspE*(pr100)–Avr3a(22–147). **d**, Co-infiltration of *Pba* expressing *recA*(pr100)–Avr3a(22–147) and AGLO delivering R3a. **e**, Co-infiltration of *Pba* expressing *dspE*(pr100)–Avr3a(22–147) and AGLO delivering R3a, yielding a hypersensitive response, indicating Avr3a fusion protein delivery into the host cell. **f**, Leaf showing (co-) infiltrations as indicated (see Supplementary Fig. 3 for details on *pelA*–Avr3a fusion results). *Pba1043* is the wild-type strain and *PbahrcC*<sup>–</sup> is T3SS-disabled. Dotted circles indicate zones of infiltration.

and EER motifs were replaced, singly and in combination, by alanine residues, and these forms were stably expressed from a virulent (homozygous for the non-recognized allele *avr3a*; ref. 4) *P. infestans* isolate. As expected, expression of native Avr3a yielded a hypersensitive response on only plants expressing R3a, whereas transformants expressing the alanine replacement forms of Avr3a were virulent (no hypersensitive response) on these plants (Fig. 2). Alanine substitutions may have a more wide-ranging effect than simply masking the side-chains at the motifs, thus more conservative amino acid substitutions KMIK-DDK (see Supplementary Information for bioinformatics analyses) were used to replace the RXLR-EER motifs, and the modified Avr3a was stably expressed in a virulent *P. infestans* isolate. As seen for the alanine replacements, the KMIK-DDK-Avr3a transformants were virulent on R3a-expressing potato (Fig. 2). All transformants, irrespective of the level of overexpression of Avr3a native, alanine and KMIK-DDK replacement forms (Supplementary Table 2), remained virulent on potato cultivar Bintje, which lacks R3a, demonstrating that transformation had not itself affected pathogenicity (transformant K9 shown in Supplementary Fig. 4).

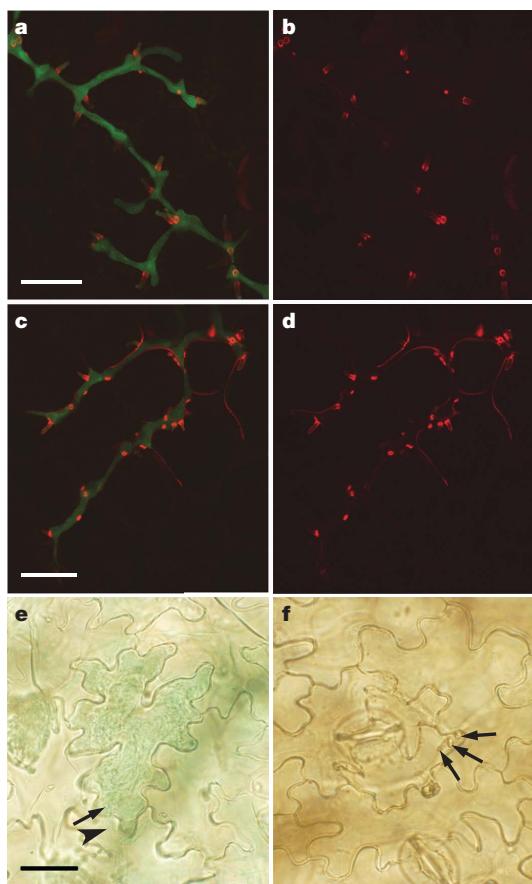
Recognition of Avr3a by R3a occurs within the plant cell<sup>4</sup> independently of the RXLR-EER motifs<sup>5</sup>. Therefore, these results suggest a role for the motifs in translocation of Avr3a into host cells. The oomycete avirulence protein ATR13 from *Hyaloperonospora parasitica*, and numerous secreted effectors from *P. infestans*, have the RXLR motif but lack, or possess a variant of, the EER motif<sup>11</sup>. Transcriptional profiling of seven such genes in *P. infestans* indicated that they were induced during infection (Supplementary Fig. 2 and Supplementary Table 1). These proteins thus either have a function outside plant cells or are able to compensate for the lack of the EER motif with alternative sequences aiding translocation.

Secretion of Avr3a was examined using *P. infestans* transformants expressing Avr3a fused to the monomeric red fluorescent protein (mRFP). Although constitutively overexpressed, localization of Avr3a–mRFP secretion was observed only from haustoria into the extra-haustorial matrix during infection of potato leaves (Fig. 3 and Supplementary Fig. 5). In contrast to similar experiments with *Plasmodium* and erythrocytes<sup>12,13</sup>, no fluorescence from



**Figure 2 | Replacement of RXLR-EER motifs with alanine residues, singly or in combination, or with amino acids KMIK-DDK, prevents delivery of Avr3a into host cells.** Leaves of potato cultivar Pentland Ace (*R3a*), infected with: **a**, *P. infestans* isolate 88069 (*avr3a*); **b**, *P. infestans* transformant K9 expressing Avr3a encoding Avr3a with RXLR-EER; **c**, transformant 5A11 expressing Avr3a encoding alanine replacement AAAA-EER; **d**, transformant 4A3 expressing Avr3a encoding alanine replacement RXLR-AAA; **e**, transformant 7A4 expressing Avr3a encoding alanine replacement AAAA-AAA; **f**, *P. infestans* transformant KD2 expressing Avr3a encoding replacement KMIK-DDK. Transformant K9 was virulent on susceptible potato cultivar Bintje (Supplementary Fig. 4). Expression levels of introduced Avr3a variants in all transformants exceeded that of *avr3a* in non-transformed isolate 88069 (Supplementary Table 2).

Avr3a-mRFP was observed within infected plant cells, presumably due to dilution within plant cells, which are many-fold larger than erythrocytes. Transformants expressing Avr3a-mRFP in which the RXLR-EER motifs were replaced by alanine residues also exhibited red fluorescence at haustoria and extra-haustorial matrices (Fig. 3 and Supplementary Fig. 5), demonstrating that these motifs neither function for targeting Avr3a to haustoria, nor act as a secretion signal. Alanine-replacement transformants expressed the fusion gene from at least half the level of that for the transformant expressing the highest level of introduced native *Avr3a-mRFP* fusion (Supplementary Table 3). Nevertheless, fluorescence observed for the alanine replacement form differed in that it was both significantly brighter and was observed surrounding plant cells containing haustoria (Fig. 3, and Supplementary Fig. 5). Our interpretation of these observations is that the alanine replacement fusion protein is secreted but no longer enters host cells, instead accumulating in the extra-haustorial matrix and overspilling into the host apoplast.



**Figure 3 | Avr3a is secreted from haustoria and is translocated into the host cell in an RXLR-EER-dependent manner.** Confocal laser scanning microscopy of *Phytophthora infestans* transformants expressing green fluorescent protein to define hyphal cytoplasm, and translational fusions of Avr3a with the monomeric red fluorescent protein. **a**, Overlay projection of signal peptide (SP)-RXLR-EER-Avr3a-mRFP in infected potato leaf tissue showing secretion of Avr3a specifically from finger-like haustoria. **b**, Red channel image of the same field as in **a**. **c**, Overlay projection of SP-AAAA-AAA-Avr3a-mRFP in infected potato, showing red fluorescence in haustoria and the host apoplast adjoining haustoria. **d**, Red channel image of the same field as in **c**. **e**, Light microscopy of *P. infestans* transformant RGUS9 expressing SP-RXLR-EER-GUS fusion (20× objective lens). An intercellular hypha (solid arrowhead) passes underneath the leaf epidermal cell exhibiting GUS activity, and penetrates this cell with a haustorium (long arrow). **f**, *P. infestans* transformant 7AGUS3 expressing SP-AAAA-AAA-GUS fusion (20× objective lens). No GUS activity can be seen in the invaded epidermal cell; at least three haustoria have penetrated this cell (long arrows). Scale bars represent 25 μm for **a** and **b**, 30 μm for **c** and **d**, and 50 μm for **e** and **f**.

To investigate further the translocation of Avr3a from haustoria into the plant cell, the signal peptide (SP)-RXLR-EER-encoding domain of *Avr3a*, and a version in which the motifs were replaced with alanine residues (SP-AAAA-AAA) were fused to the amino terminus of the *Escherichia coli* *gusA* gene, which encodes β-glucuronidase (GUS), and stably expressed in *P. infestans*. *GusA* was chosen because its product is active within plant cells but inactive in the apoplast<sup>19</sup>, and is thus an ideal reporter for translocation of proteins to the inside of plant cells. SP-RXLR-EER-GUS and SP-AAAA-AAA-GUS transformants were selected that expressed and secreted high levels of active GUS into culture medium (Supplementary Fig. 6). However, when infecting potato, GUS activity was observed only with the SP-RXLR-EER-GUS transformants, and only within plant cells in contact with haustoria (Fig. 3 and Supplementary Fig. 7). No GUS activity was observed within plant cells or in the apoplast in the case of the SP-AAAA-AAA-GUS transformants (Fig. 3 and Supplementary Fig. 7), indicating that GUS was not translocated into the plant cell.

Our results show that, following secretion from haustoria into the extra-haustorial matrix, Avr3a is translocated into host cells by an unknown mechanism dependent on the RXLR-EER motifs. This dependence implies that other *P. infestans* RXLR-EER class effectors are transported by a similar mechanism. The similarities of function and positioning between this oomycete translocation signal and the malarial host targeting signal RXLXE/D/Q within effector proteins have been considered to reinforce the phylogenetic relatedness of these organisms<sup>15</sup>. However, haustorium-forming and non-haustorium-forming fungal plant pathogens, which do not possess RXLR-EER class effectors, also secrete avirulence and effector proteins that are perceived by cytoplasmic host resistance proteins, although their mode of translocation is unknown. It is possible that they also contain peptide motifs required for translocation into host cells.

The bacterial T3SS is a well-characterized mechanism for delivery of effectors from bacteria into host cells. However, identification of a protein sequence motif for T3SS-dependent translocation and, consequently, predicting the full range of T3SS-delivered proteins from plant pathogens such as *Pseudomonas syringae*, have provided many challenges<sup>20</sup>. A functionally analogous mechanism for delivery of oomycete effectors into host cells has yet to be elucidated. Indeed, it is possible that, when RXLR-EER class effectors are secreted into the extra-haustorial matrix, and thus in intimate contact with the host cell membrane, that these motifs alone effect translocation, potentially using host endocytic mechanisms. Nevertheless, determination of the RXLR-EER translocation signal permits facile identification of this class of potential translocated virulence proteins. This discovery will facilitate investigations into how this important group of pathogens manipulate host defences and metabolism to establish disease. Moreover, comparisons between the RXLR-EER effector complements of sequenced oomycete genomes, and studies of co-evolution of effectors and their host target proteins, will yield insights into how host range is defined.

## METHODS SUMMARY

**Gene expression.** Plant growth, manipulation of *P. infestans* isolate 88069, plant inoculation, RNA extraction, complementary DNA synthesis, and real-time RT-PCR were carried out as described<sup>21,22</sup> (Supplementary Table 4).

**Bioinformatics.** Hidden Markov models and an extended heuristic method<sup>14,23</sup> were both used to predict secreted RXLR-EER class effectors in the draft *P. infestans* genome (see Supplementary Information bioinformatics analyses).

**Pectobacterium atrosepticum delivery of Avr3a.** The native promoter and region encoding approximately the 100 N-terminal amino acids of genes *dspE*, *recA* or *pelA* were fused to *Avr3a* lacking a signal peptide sequence, and transformed into *P. atrosepticum*. Fully expanded leaves of *Nicotiana benthamiana* plants were infiltrated with each *P. atrosepticum* strain.

**P. infestans transformation.** *Avr3a*, with and without the RXLR and/or EER replaced by alanine or KMIK-DDK residues, was cloned into the oomycete expression vector pTOR<sup>24</sup>, incorporating mismatches across the RXLR or EER



motifs. Avr3a and alanine replacement variants were fused in frame to the *mRFP* gene in vector pTORmRFP4. Plasmid pVWII for GFP expression has been described previously<sup>25</sup>. Avr3a SP-RXLR-EER- or SP-AAAA-AAA-encoding sequences were fused in frame to the *gusA* gene in pTOR. Transformation of *P. infestans* was as described<sup>26</sup>, modified as described in Supplementary Methods. ***P. infestans* potato inoculation.** Sporangia were inoculated onto detached leaflets of potato cultivars Pentland Ace (*R3a*) or Bintje (*R*-gene-free), and incubated at 20 °C for 5 d to enable disease symptoms to develop. Untransformed isolate 88069 was used as a control.

**Confocal microscopy.** GFP (488 nm excitation; emission 500–530 nm) and mRFP (561 nm excitation; emission 600–630 nm) were imaged on a Leica TCS-SP2 AOBs confocal microscope.

**Histochemical GUS assay.** Infected leaf tissue of potato cv. Bintje was submerged in assay buffer containing GUS substrate for 24 h at 20 °C. Leaf tissue was cleared with 70:30 ethanol:water before microscopy on an Olympus BH-2 microscope.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 11 July; accepted 24 August 2007.

Published online 30 September 2007.

- Alfano, J. R. & Collmer, A. Type III secretion system effector proteins: double agents in bacterial disease and plant defense. *Annu. Rev. Phytopathol.* **42**, 385–414 (2004).
- Birch, P. R. J., Rehmany, A. P., Pritchard, L., Kamoun, S. & Beynon, J. L. Trafficking arms: oomycete effectors enter host plant cells. *Trends Microbiol.* **14**, 8–11 (2006).
- Ellis, J., Catanzariti, A. M. & Dodds, P. The problem of how fungal and oomycete avirulence proteins enter plant cells. *Trends Plant Sci.* **11**, 61–63 (2006).
- Armstrong, M. R. *et al.* An ancestral oomycete locus contains late blight avirulence gene *Avr3a*, encoding a protein that is recognised in the host cytoplasm. *Proc. Natl Acad. Sci. USA* **102**, 7766–7771 (2005).
- Bos, J. I. *et al.* The C-terminal half of *Phytophthora infestans* RXLR effector AVR3a is sufficient to trigger R3a-mediated hypersensitivity and suppress INF1-induced cell death in *Nicotiana benthamiana*. *Plant J.* **48**, 165–176 (2006).
- Jia, J., Adams, S. A., Bryan, G. T., Hershey, H. P. & Valent, B. Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *EMBO J.* **19**, 4004–4014 (2000).
- Dodds, P. N., Lawrence, G. J., Catanzariti, A. M., Ayliffe, M. A. & Ellis, J. G. The *Melampsora lini* AvrL567 avirulence genes are expressed in haustoria and their products are recognized inside plant cells. *Plant Cell* **16**, 755–768 (2004).
- Ridout, C. J. *et al.* Multiple avirulence paralogues in cereal powdery mildew fungi may contribute to parasite fitness and defeat of plant resistance. *Plant Cell* **18**, 2402–2414 (2006).
- Allen, R. L. *et al.* Host–parasite co-evolutionary conflict between *Arabidopsis* and downy mildew. *Science* **306**, 1957–1960 (2004).
- Shan, W. *et al.* The *Avr1b* locus of *Phytophthora sojae* encodes an elicitor and a regulator required for avirulence on soybean plants carrying resistance gene *Rps1b*. *Mol. Plant Microbe Interact.* **17**, 394–403 (2004).
- Rehmany, A. P. *et al.* Differential recognition of highly divergent downy mildew avirulence gene alleles by *RPP1* resistance genes from two *Arabidopsis* lines. *Plant Cell* **17**, 1839–1850 (2005).
- Hiller, N. L. *et al.* A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science* **306**, 1934–1937 (2004).
- Marti, M., Good, R. T., Rug, M., Kuepfer, E. & Cowman, A. F. Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science* **306**, 1930–1933 (2004).
- Bhattacharjee, S. *et al.* The malarial host-targeting signal is conserved in the Irish potato famine pathogen. *PLoS Pathogens* **2**, e50, doi:10.1371/journal.ppat.0020050 (2006).
- Haldar, K., Kamoun, S., Hiller, N. L., Bhattacharjee, S. & van Ooij, C. Common infection strategies of pathogenic eukaryotes. *Nature Rev. Microbiol.* **4**, 922–931 (2006).
- Tyler, B. M. *et al.* *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* **313**, 1261–1266 (2006).
- Toth, I. K. & Birch, P. R. J. Rotting softly and stealthily. *Curr. Opin. Plant Biol.* **8**, 424–429 (2005).
- Holeva, M. C. *et al.* Use of a pooled transposon mutation grid to demonstrate roles in disease development for *Erwinia carotovora* subsp. *atroseptica* putative Type III secreted effector (DspE/A) and helper (HrpN) proteins. *Mol. Plant Microbe Interact.* **17**, 943–950 (2004).
- Denecke, J., Botterman, J. & Deblair, R. Protein secretion in plant cells can occur via a default pathway. *Plant Cell* **2**, 51–59 (1990).
- Lindeberg, M. *et al.* Closing the circle on the discovery of genes encoding Hrp regulon members and type III secretion system effectors in the genomes of three model *Pseudomonas syringae* strains. *Mol. Plant Microbe Interact.* **19**, 1151–1158 (2006).
- Grenville-Briggs, L. J. *et al.* Elevated amino acid biosynthesis in *Phytophthora infestans* during appressorium formation and potato infection. *Fungal Genet. Biol.* **42**, 244–256 (2005).
- Avrova, A. O., Venter, E., Birch, P. R. J. & Whisson, S. C. Profiling and quantifying differential gene transcription in *Phytophthora infestans* prior to and during the early stages of potato infection. *Fungal Genet. Biol.* **40**, 4–14 (2003).
- Torto, T. A. *et al.* EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen *Phytophthora*. *Genom. Res.* **13**, 1675–1685 (2003).
- Blanco, F. A. & Judelson, H. S. A bZIP transcription factor from *Phytophthora* interacts with a protein kinase and is required for zoospore motility and plant infection. *Mol. Microbiol.* **56**, 638–648 (2005).
- van West, P. *et al.* Green fluorescent protein (GFP) as a reporter gene for the plant pathogenic oomycete *Phytophthora palmivora*. *FEMS Microbiol. Lett.* **178**, 71–80 (1999).
- Judelson, H. S., Tyler, B. M. & Michelmore, R. W. Transformation of the oomycete pathogen, *Phytophthora infestans*. *Mol. Plant Microbe Interact.* **4**, 602–607 (1991).
- Guttman, D. S. *et al.* A functional screen for the Type III (Hrp) secretome of the plant pathogen *Pseudomonas syringae*. *Science* **295**, 1722–1726 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature). A summary figure is also included.

**Acknowledgements** This work was supported by grants from the Scottish Executive Environment and Rural Affairs Department (SEERAD) to S.C.W., P.R.J.B. and I.K.T. J.G.M. is supported by the Universidad Nacional de Colombia sede Palmira Agricultural Sciences department and the European Union Alban Programme. P.v.W. is supported by The Royal Society; L.M. is supported by a Commonwealth Scholarship and Fellowship Plan. The authors thank G. Cowan, H. Liu and E. Venter for technical assistance.

**Author Contributions** S.C.W., P.R.J.B., P.v.W. and I.K.T. developed the concept and designed experiments. S.C.W. and S.G. performed *P. infestans* transformations and plant inoculations. P.C.B. carried out confocal microscopy and advised on cell biology. S.C.W. performed GUS assays and light microscopy. A.O.A. and J.G.M. quantified gene expression. Antibody detection of tagged transformants was performed by I.H. and S.C. L.M., J.G.M., E.M.G. and M.R.A. carried out experiments with *P. atroscopicum*. L.P. conducted all bioinformatics analyses.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to S.C.W. (Steve.Whisson@scri.ac.uk) or P.R.J.B. (Paul.Birch@scri.ac.uk).



## METHODS

**RNA extraction and real-time RT-PCR.** Growth of potato plants, *P. infestans* isolate 88069, and plant inoculation were carried out as described<sup>21</sup>. Life cycle stages of *P. infestans* (axenic cultured mycelium, sporangia, zoospores, germinated cysts) were also prepared as described<sup>21</sup>. RNA extraction, first strand complementary DNA synthesis and SYBR green real time RT-PCR assays were carried out as described<sup>22</sup>. Primer sequences for real-time RT-PCR are listed in Supplementary Table 4.

**Bioinformatic identification of candidate RXLR-EER sequences.** A heuristic method to identify sequences containing a signal peptide and the RXLR motif<sup>14,23</sup> was extended to incorporate the presence of an [ED][ED][KR] motif downstream and within 40 amino acids of the RXLR motif. A hidden Markov model was also constructed, based on an alignment of the RXLR-EER-containing regions of 38 proteins from *P. infestans* that are known to be induced during infection. Each model was used to compile a list of candidate open reading frames that potentially encode secreted RXLR-EER class effectors in the draft *P. infestans* genome (described in detail in Supplementary Information bioinformatics analyses).

**Expression of Avr3a fusion constructs in *Pectobacterium atrosepticum*.** The native promoter and the region encoding approximately 100 amino acids of the N terminus of genes *dspE*, *recA* or *pelA* were PCR-amplified using primer gene\_F and primer gene\_Avr3a\_R from *P. atrosepticum* (Supplementary Table 5). *Avr3a*(67–441) (nucleotides 67 to 441; GenBank AJ893357; ref. 4) (lacking signal-peptide-encoding sequences) was PCR-amplified from pRTL2-Avr3a (ref. 4) using primer Avr3a\_gene\_F and Avr3a\_R (Supplementary Table 5). N-terminus-encoding PCR products were fused to *Avr3a* purified products by extension followed by overlapping PCR conducted with gene\_F and Avr3a\_R primers. *PelA-Flag-Avr3a* fusions were constructed with sequence encoding a Flag epitope tag located in the PCR overlap between *pelA* and *Avr3a*. Products were purified and cloned into pGEM-T Easy vector (Promega), then electroporated into *Escherichia coli* DH10B cells (Invitrogen). Plasmid preparations were sequenced to confirm the fidelity of the *P. atrosepticum* gene-*Avr3a* fusions. Plasmids were then electroporated into *P. atrosepticum* strain SCRI1043. A full description of transformation constructs and protocol can be found in the Supplementary Methods.

For hypersensitive response assays, fully expanded leaves of *Nicotiana benthamiana* plants (approximately 6 weeks old) were infiltrated on the lower surface by first stabbing lightly with a needle then infiltrating with a blunt syringe. Approximately 100 µl of each strain, *P. atrosepticum* wild-type strain SCRI1043 or *hrcC*<sup>−</sup> mutants containing plasmids with *P. atrosepticum* gene-*Avr3a* fusions, with *A. tumefaciens* strain AGLO carrying pBINplus-R3a were infiltrated singly or co-infiltrated as described<sup>4</sup>. All strains were grown overnight in LB broth with an appropriate antibiotic to  $2 \times 10^8$  colony forming units ml<sup>−1</sup> then mixed at a ratio of 2:1 *A. tumefaciens*:*P. atrosepticum*. Plants were kept at 18 °C at high humidity overnight (16 h). Images were captured at 120 h using a Nikon digital camera.

***P. infestans* transformation vector construction and transformation.** The *Avr3a* open reading frame was PCR amplified (primers Avr3F and Avr3R; Supplementary Table 5) and cloned into the oomycete constitutive expression vector pTOR (ref. 24). *Avr3a*, with sequence encoding the RXLR or EER to be replaced by alanine residues, was generated by amplification of the *Avr3a* gene in two segments (primer combinations Avr3ClaF2 and 5ANgoM4R or 4ANotR; primers 5ANgoM4F, 4ANotF, or KMIKDDKNgoF and Avr3aStopSacR), incorporating mismatches across the RXLR or EER motifs. The two segments

were then ligated together and inserted into the pTOR vector. *Avr3a* with sequence encoding both RXLR and EER motifs to be replaced by alanine residues was generated by starting with the EER alanine replacement sequence and proceeding through the same strategy used to generate the RXLR to AAAA sequence. The *mRFP* gene was PCR amplified (primers mRFPBsiF and mRFPSacR) and incorporated into the pTOR vector to yield pTORMRFP4. *Avr3a* and the RXLR alanine replacement variants described previously were then cloned in frame without a stop codon (primers Avr3ClaF2 and Avr3BsiR), ahead of the mRFP gene. Insert integrity and correct reading frame orientation of the cloned inserts were verified by sequencing. Plasmid pVWII for enhanced green fluorescent protein expression has been described previously<sup>25</sup>.

RXLR-EER-GUS and AAAA-AAA-GUS fusions were constructed by amplification of the *Avr3a* SP-RXLR-EER-encoding sequence (primers Avr3ClaF2 and RXLRnotR), SP-AAAA-AAA-encoding sequence (primers Avr3ClaF2 and 7AnotR), and *gusA* (primers GUSnotF and GUSsacR), followed by ligation of the SP-RXLR-EER- or SP-AAAA-AAA-encoding sequence to the *gusA* gene in pTOR. Transformation of *E. coli* and determination of construct integrity were as described.

Stable transformation of *P. infestans* was achieved using a PEG-CaCl<sub>2</sub>-Lipofectin protocol<sup>26</sup> (<http://138.23.152.128/protocols/protocols.html>), modified by substituting Novozym 234 with a mixture of 5 mg ml<sup>−1</sup> lysing enzymes (from *Trichoderma harzianum*; Sigma L1412) and 2 mg ml<sup>−1</sup> cellulase (from *Trichoderma reesei*; Sigma C8546) to release protoplasts from germinating sporangia. A full description of transformation constructs and protocol can be found in Supplementary Methods.

**Inoculation of *P. infestans* transformants onto potato leaves.** Sporangia ( $5 \times 10^4$  ml<sup>−1</sup>) of stable transformants were inoculated in 10 µl droplets onto detached leaflets of potato cultivars Pentland Ace (*R3a*) or Bintje (*R*-gene-free). Inoculated leaflets were incubated at 20 °C for 5 d to enable disease symptoms to develop. For each transformant, six leaflets were inoculated on three independent occasions. Samples were taken at 120 h.p.i. for RNA extraction and real-time RT-PCR analysis, as described previously. As a control, untransformed isolate 88069 was used. Infected leaves were photographed under polarized light to avoid reflection from the leaf surfaces.

**Imaging and quantification of fluorescent proteins.** All GFP and mRFP imaging was conducted on a Leica TCS-SP2 AOBs confocal microscope using HC PL FLUOTAR ×63 0.9 (numerical aperture) (as for Fig. 3a–d), HCX APO L U-V-I ×40 0.8 or HCX APO L U-V-I ×20 0.5 water-dipping lenses. For all images shown in this paper, the 561 nm line from the 'lime' diode laser was set at 100% on the AOBs. PMT2 was set at either 670 V (Fig. 3a, b) or 603 V (Fig. 3c, d). The total mRFP fluorescence of each haustorium with a circular profile was measured within a region-of-interest with an area of approximately 13 µm<sup>2</sup>, using the Leica confocal software quantification package. Fluorescence measurements were exported to Microsoft Excel and standard errors of the mean calculated.

**Histochemical GUS assay.** GUS substrate (0.3 mg ml<sup>−1</sup> 5-bromo-4-chloro-3-indolyl β-D-glucuronide) was allowed to infiltrate into infected leaf tissue of potato cv. Bintje submerged in assay buffer (50 mM NaPO<sub>4</sub>, pH 7.0, 0.5% Triton X-100) for 24 h at 20 °C. This gentle procedure was adopted because vacuum infiltration of substrate and 37 °C incubation resulted in disruption of host and pathogen cells, giving non-specific localization of GUS activity. Leaf tissue was cleared with 70:30 ethanol:water before microscopy on an Olympus BH-2 microscope fitted with 6.3×, 10×, 20× or 40× objective lenses, and an Olympus DP70 digital camera for image capture.

# Histone demethylase JHDM2A is critical for *Tnp1* and *Prm1* transcription and spermatogenesis

Yuki Okada<sup>1,2</sup>, Greg Scott<sup>3</sup>, Manas K. Ray<sup>3</sup>, Yuji Mishina<sup>3,4</sup> & Yi Zhang<sup>1,2</sup>

Recent studies indicate that, similar to other covalent modifications, histone lysine methylation is subject to enzyme-catalysed reversal<sup>1,2</sup>. So far, LSD1 (also known as AOF2) and the jumonji C (JmjC)-domain-containing proteins have been shown to possess histone demethylase activity. LSD1 catalyses removal of H3K4me2/H3K4me1 through a flavin-adenine-dinucleotide-dependent oxidation reaction<sup>3</sup>. In contrast, JmjC-domain-containing proteins remove methyl groups from histones through a hydroxylation reaction that requires  $\alpha$ -ketoglutarate and Fe(II) as cofactors<sup>4</sup>. Although an increasing number of histone demethylases have been identified and biochemically characterized<sup>1,2</sup>, their biological functions, particularly in the context of an animal model, are poorly characterized. Here we use a loss-of-function approach to demonstrate that the mouse H3K9me2/1-specific demethylase JHDM2A (JmjC-domain-containing histone demethylase 2A, also known as JMJD1A) is essential for spermatogenesis. We show that *Jhdm2a*-deficient mice exhibit post-meiotic chromatin condensation defects, and that JHDM2A directly binds to and controls the expression of transition nuclear protein 1 (*Tnp1*) and protamine 1 (*Prm1*) genes, the products of which are required for packaging and condensation of sperm chromatin. Thus, our work uncovers a role for JHDM2A in spermatogenesis and reveals transition nuclear protein and protamine genes as direct targets of JHDM2A.

We have previously demonstrated that the histone demethylase JHDM2A has an important role in transcriptional activation mediated by the androgen receptor<sup>5</sup>. To further explore the biological function of this protein, we examined its expression pattern in mice. Similar to a previous report<sup>6</sup>, we found that *Jhdm2a* is highly expressed in testis at both the messenger RNA and protein levels (data not shown). Further analysis of *Jhdm2a* by polymerase chain reaction with reverse transcription (RT-PCR) indicated that *Jhdm2a* levels (Fig. 1a, filled bars) increased significantly during spermatogenesis (~70-fold increase from day 7 testis to day 30 testis). In contrast, the level of the highly related gene *Jhdm2b* (also known as *Jmjd1b*) only varied 0.5–2.5-fold during spermatogenesis (Fig. 1a, open bars). When compared with several spermatogenesis marker genes (namely, zinc-finger- and BTB-domain-containing 16 (*Plzf*, also known as *Zbtb16*), synaptonemal complex protein 2 (*Sycp2*), transition nuclear protein 1 (*Tnp1*) and protamine 2 (*Prm2*)), *Jhdm2a* seems to be expressed during and after meiosis, preceding the expression of transition proteins and protamines (Fig. 1a, bottom panels). Western blot analysis using protein extracts of partially purified spermatocytes, round spermatids and elongating spermatids indicated that the level of JHDM2A is at the highest in round spermatids (Fig. 1b). Immunostaining revealed that JHDM2A protein is expressed in various spermatogenic stages of seminiferous tubules (Supplementary Fig. 1a), particularly in the nuclei of round spermatids (Fig. 1c, d), as well as in the nuclei and cytoplasm of Sertoli cells (Fig. 1c, arrowheads). Notably, JHDM2A

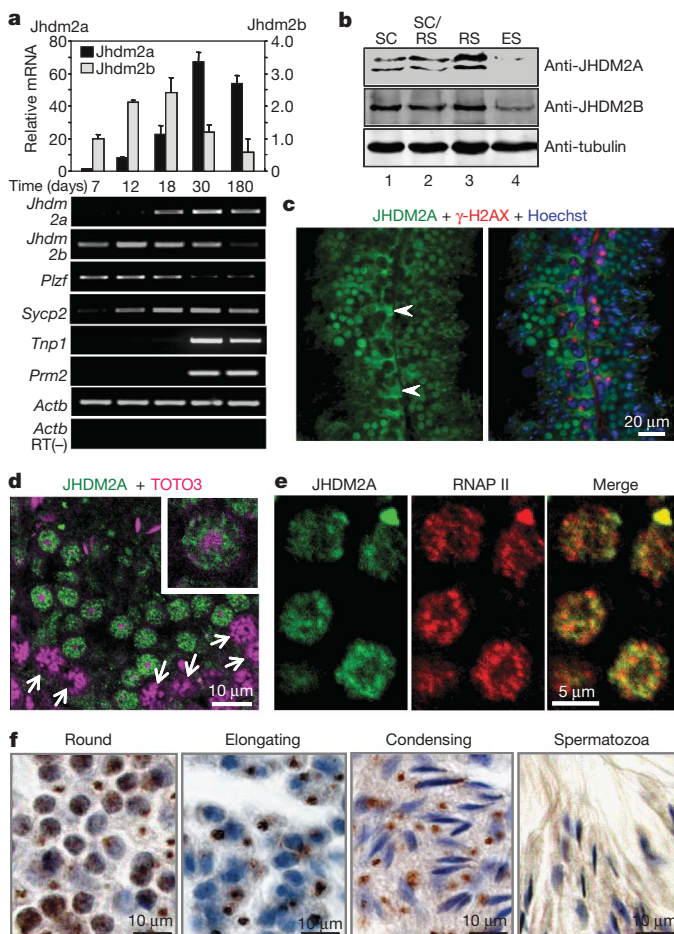
protein was not detectable in spermatogonia or  $\gamma$ -H2AX-positive (also known as  $\gamma$ -H2AFX-positive) early pachytene spermatocytes (Fig. 1c, d arrows). A detailed analysis indicated that JHDM2A protein starts to be detected in the late pachytene stage, increases in diplotene and secondary spermatocytes, and reaches its highest level in round spermatids (Supplementary Fig. 1b, c). Interestingly, JHDM2A co-localizes with RNA polymerase II (phospho-Ser5) in round spermatids as nuclear dots (Fig. 1e), and is excluded from the chromocentre and perinuclear heterochromatin (Fig. 1d, inset); this is consistent with the notion that JHDM2A functions as a transcriptional co-activator<sup>5</sup>. When spermatids start to elongate, JHDM2A localizes to the cytoplasm and forms distinct foci, which persist until spermiogenic step 13 and disappear in mature spermatozoa (Fig. 1f and Supplementary Fig. 1c).

The dynamic nature of JHDM2A expression during spermiogenesis indicates that it may have an important role in the late stages of male-germ-cell development. To explore this possibility, the mouse *Jhdm2a* gene was disrupted by insertion of a  $\beta$ -galactosidase/neomycin ( $\beta$ -Geo) cassette in intron 10 (or intron 7 of splicing variant 2) (Fig. 2a and Supplementary Fig. 2a). Chimaeras and heterozygous (*Jhdm2a*<sup>+/-G</sup>) mice were obtained using standard procedures. Homozygous mice that carry the  $\beta$ -Geo alleles (*Jhdm2a*<sup>G/G</sup>) were obtained by crossing heterozygous pairs. Homozygous mice, verified by Southern blot analysis (Supplementary Fig. 2b, c), were viable and were born at mendelian ratios (data not shown). Expression of the *Jhdm2a*(exon 1–10)- $\beta$ -Geo fusion transcript was confirmed by RT-PCR in *Jhdm2a*<sup>+/-G</sup> and *Jhdm2a*<sup>G/G</sup> mice (Fig. 2b, bottom panel). However, a low level (~10%) of wild-type mRNA was also detected in *Jhdm2a*<sup>G/G</sup> mice (Fig. 2b, top panel, lane 5, and Fig. 2c).

The *Jhdm2a* mutant allele encodes the first 506 amino acids of JHDM2A and is fused to  $\beta$ -GEO (Fig. 2d). Given that the catalytic JmjC domain is deleted, the mutant protein will not be enzymatically active. Whole-mount  $\beta$ -Gal staining confirmed the expression of the  $\beta$ -Gal protein in testis in a dose-dependent manner (Fig. 2e). Western blot analysis demonstrated that the JHDM2A protein in the testis was detected as a doublet, which corresponds to the full-length (V1, 152 kilodaltons (kDa)) and the splicing variant (V2, 139 kDa) (Fig. 2f, top panel, and Supplementary Fig. 2a). Although the splicing variant was not detectable, lower levels of full-length JHDM2A protein still remain in the *Jhdm2a*<sup>G/G</sup> mice (Fig. 2f, top panel). As expected, the JHDM2A(1–506)- $\beta$ -Gal fusion protein was produced in both *Jhdm2a*<sup>+/-G</sup> and *Jhdm2a*<sup>G/G</sup> mice (Fig. 2f, second panel), and this fusion protein was also detectable by an anti-JHDM2A antibody (data not shown). Neither JHDM2B, a related H3K9 demethylase<sup>5</sup>, nor DNMT3B, a *de novo* DNA methyltransferase expressed in early meiosis and round spermatids<sup>7</sup>, was affected by *Jhdm2a* disruption (Fig. 2f).

<sup>1</sup>Howard Hughes Medical Institute, <sup>2</sup>Department of Biochemistry and Biophysics, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7295, USA. <sup>3</sup>The Knock Out Core, <sup>4</sup>Molecular Developmental Biology Group, Laboratory of Reproductive and Developmental Toxicology, National Institutes of Health, Research Triangle Park, North Carolina 27709, USA.

Consistent with having smaller testes (Fig. 2e and Supplementary Table 1) and a significantly lower sperm count (Supplementary Table 1), *Jhdm2a*<sup>G/G</sup> male mice were infertile (data not shown). Histological analysis revealed multiple vacuolations in seminiferous tubules and fewer elongated spermatids in *Jhdm2a*<sup>G/G</sup> testis (Supplementary Fig. 3a). A detailed examination revealed that noticeable morphological defects start from the spermiogenic stage XI–XII spermatids having less-elongated heads (Fig. 3a, inset). Spermatids in *Jhdm2a*<sup>G/G</sup> mice remain unelongated at stage I and II–III (Fig. 3a, insets); the number of spermatids greatly decreased and the spermatids are not properly elongated at stage IV–VIII (Fig. 3a). At stage VIII, multiple spermatids sharing their cytoplasm and having multiple tails are observed

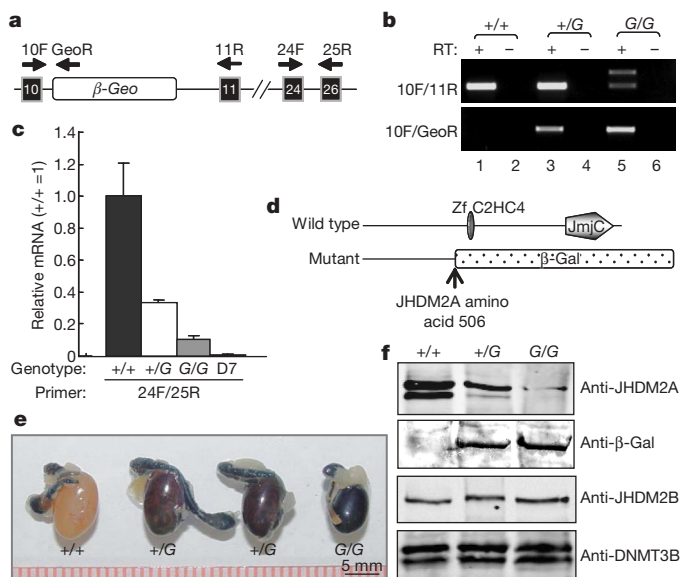


**Figure 1** | *Jhdm2a* expression is largely restricted to post-meiotic male germ cells. **a**, Top panel, quantitative RT–PCR analysis of the *Jhdm2a* and *Jhdm2b* expression levels in testis from mice of different ages. The expression level in day 7 mouse testis was arbitrarily set as 1. Data presented are means  $\pm$  s.d. from three independent experiments. Lower panels are RT–PCR results of various genes in testes of different developmental stages. *Plzf*, spermatogonia-specific; *Sycp2*, spermatocyte-specific; *Tnp1*, post-meiosis-specific; and *Prm2*, post-meiosis-specific. *Actb* ( $\beta$ -actin) serves as a control. RT–, reverse transcriptase minus. **b**, Western blot analysis of JHDM2A and JHDM2B expression. ES, elongating spermatid; RS, round spermatid; and SC, spermatocyte. Lane 1, SCs with  $\sim$ 30% RS contamination. Lane 2, SCs and RSs ( $\sim$ 50% each). Lane 3, RSs with  $\sim$ 30% SC contamination. Lane 4, mouse ES cells. Tubulin serves as a loading control. **c**, Expression of JHDM2A in seminiferous tubules. JHDM2A (green),  $\gamma$ -H2AX (red) and Hoechst (DNA, blue). Arrowheads, Sertoli cells. **d**, Confocal microscopic analysis of JHDM2A in stage V seminiferous tubules. JHDM2A (green) and TOTO3 (DNA, magenta). Arrows, pachytene spermatocytes. **e**, Co-localization of JHDM2A (green) and RNA polymerase II (RNAP II, red) in round spermatids. **f**, Dynamic expression pattern of JHDM2A in various stages of post-meiotic male germ cells. Immunopositive cells are stained brown. Counterstaining by haematoxylin is blue.

(Fig. 3a, stage VIII, arrow). Electron microscopic analysis revealed abnormal nuclear structure including loss of chromocentre organization (Fig. 3b, step 2–3) and loss of polarity of chromatin distribution (Fig. 3b, steps 7 and 9). Although the acrosomal structure in round spermatids seems to be normal at step 7, it is not fully formed in step 13 (Fig. 3b). In contrast to the multiple post-meiotic defects, pre-meiotic germ cells, Sertoli cells and Leydig cells are morphologically normal (Supplementary Fig. 3a, b). Thus, the main function of JHDM2A is in the post-meiotic stage of spermatogenesis.

In agreement with the defects described above, few mature sperm were observed in the epididymis of *Jhdm2a*<sup>G/G</sup> mice (Supplementary Fig. 3c). Of the few sperm recovered from the epididymis of the mutant mice, all have abnormally shaped heads (Fig. 3c). Although the mutant sperm have tails, 98.5% of them are immotile (Fig. 3d). To assess the packaging state of the sperm DNA, we performed acridine orange staining, an assay that can distinguish the packaging state of the sperm DNA on the basis of whether the sperm are fluorescent green only or both green and red<sup>8</sup>. The fact that the abnormal sperm heads from *Jhdm2a*<sup>G/G</sup> stained both green and red by acridine orange (Fig. 3e) indicates that the sperm heads from mutant mice are defective in chromatin condensation. Flow cytometry analysis also revealed that sperm from *Jhdm2a*<sup>G/G</sup> mice have a DNA content that peaks between round and condensed spermatids (Supplementary Fig. 3d), consistent with defects in chromatin condensation<sup>9</sup>. Collectively, the above results support the idea that incomplete chromatin condensation is the cause of the infertility of male *Jhdm2a*<sup>G/G</sup> mice.

Mice lacking the H3K9 methyltransferases SUV39H1 and SUV39H2 have meiotic defects<sup>10</sup>, indicating that maintaining an appropriate H3K9 methylation level is important for meiosis during spermatogenesis. To evaluate whether loss of *Jhdm2a* function affects



**Figure 2** | *Jhdm2a* mutant mice are hypomorphic. **a**, Diagram of the *Jhdm2a* mutant allele derived from clone YHA186 (BayGenomics). Arrows, primers used in **b** and **c**. Black boxes with numbers refer to the respective exons of the *Jhdm2a* gene. **b**, RT–PCR analysis of the expression of the wild-type and mutant alleles in *Jhdm2a*<sup>+/+</sup>, *Jhdm2a*<sup>+/G</sup> and *Jhdm2a*<sup>G/G</sup> mouse testis. Total RNA isolated from testis was used for RT–PCR. **c**, Quantitative RT–PCR analysis of the expression of the 3' end of *Jhdm2a*. D7, total RNA isolated from wild-type day 7 mouse testis. The expression level in adult *Jhdm2a*<sup>+/+</sup> is arbitrarily defined as 1. Data presented are means  $\pm$  s.d. from three mice for each genotype. **d**, Diagram of the wild-type and mutant protein alleles. The mutant allele creates a fusion protein of JHDM2A(1–506)– $\beta$ -Gal. Zf C2HC4, C2HC4 type zinc finger motif. **e**, Whole-mount  $\beta$ -galactosidase staining in testes of indicated genotypes. **f**, Western blot analysis of JHDM2A and JHDM2A(1–506)– $\beta$ -Gal fusion protein expression in testis of indicated genotypes. Equal amounts of total testis nuclear extracts were used in the blot. The antibodies used are indicated.



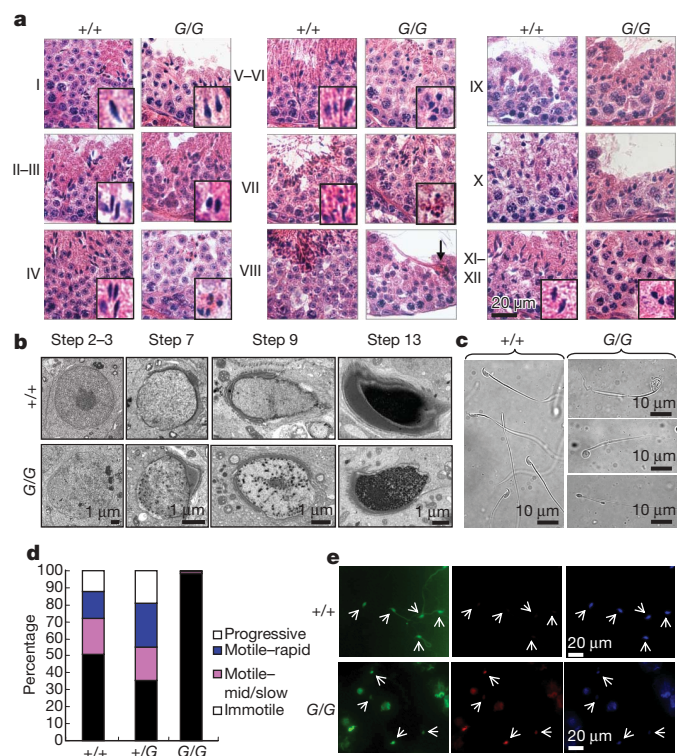
global H3K9 methylation levels during spermatogenesis, we performed immunostaining. In contrast to a previous report<sup>11</sup>, we found that both H3K9me2 and H3K9me3 levels were relatively low in spermatocytes, and were high in spermatogonia and spermatids in wild-type testis (Supplementary Fig. 4a, b). This methylation pattern is maintained in the *Jhdm2a*<sup>G/G</sup> seminiferous tubules (Supplementary Fig. 4a, b). However, the H3K9me3 staining pattern in round spermatids became more diffuse in the mutant compared to in the wild-type testis (Supplementary Fig. 4d), although no significant change in the total H3K9 methylation levels was detected (Supplementary Fig. 4e). The decreased H4 acetylation level was probably due to a decreased number of elongating/condensing spermatids in the mutant testis (Supplementary Fig. 4c, e)<sup>12</sup>. Because active DNA demethylation has been reported in spermatocytes<sup>13</sup>, and there is a tight correlation between DNA methylation and H3K9 methylation during germ-cell development<sup>14</sup>, we analysed the effect of loss of *Jhdm2a* function on DNA methylation in seminiferous tubules. Results shown in Supplementary Fig. 5 revealed no significant differences. Therefore, loss of *Jhdm2a* function does not have global effect on H3K9 or DNA methylation in testis.

Multiple molecular events have to occur for a round spermatid to become a mature sperm. These events include chromatin

condensation, reorganization of the spermatid nucleus, formation of an acrosome and assembly of a sperm tail<sup>15</sup>. To accomplish these events, a number of postmeiotic proteins including transition nuclear proteins, protamines and testis-specific histone variants, such as H1T (also known as histone cluster H1T, HISTH1T), H1T2 (also known as histone family member N, testis specific, H1FNT) and HILS1 (histone linker H1 domain, spermatid-specific 1), have to be synthesized properly. Consequently, mice that lack the proteins mentioned above display impaired spermiogenesis and male infertility<sup>16</sup>. Defective chromatin condensation similar to that observed in the *Jhdm2a*<sup>G/G</sup> mice was reported in the knockout mice for transition nuclear proteins, protamines, H1T2 and TLP (TBP-like protein, also known as TBPL1), raising the possibility that *Jhdm2a* may contribute to spermiogenesis by regulating expression of some of these genes<sup>17–20</sup>. To explore this possibility, we attempted to purify round spermatids, in which JHDM2A is highly expressed, from *Jhdm2a*<sup>+/+</sup> and *Jhdm2a*<sup>G/G</sup> testes. Purity of the isolated cells was confirmed by examining the expression profile of stage-specific genes (Supplementary Fig. 6). Quantitative (q)RT–PCR demonstrated that, although the expression of most genes tested was not significantly altered, the expression of *Tnp1* and *Prm1* was significantly reduced in the *Jhdm2a*<sup>G/G</sup> round spermatids (Fig. 4a).

Chromatin immunoprecipitation (ChIP) analysis using testicular cells demonstrated that JHDM2A was specifically recruited to region C of the *Prm1* gene (Fig. 4b, compare lanes 3 and 6). Although the JHDM2A(1–506)–β-Gal fusion protein can be recognized by our anti-JHDM2A antibody, the fusion protein does not occupy the *Prm1* promoter, suggesting the deleted carboxy-terminal region of JHDM2A is required for chromatin association. ChIP–qPCR analysis using purified round spermatids confirmed specific enrichment of JHDM2A to the core promoter of the *Prm1* gene (Fig. 4c). Similar results were obtained for the *Tnp1* gene, but not the *Tnp2* gene (Fig. 4c), consistent with its specific effects on *Prm1* and *Tnp1* gene expression (Fig. 4a). To evaluate the effect of JHDM2A binding on H3K9 methylation, we compared the H3K9 methylation levels of *Prm1* and *Tnp1* in *Jhdm2a*<sup>+/+</sup> and *Jhdm2a*<sup>G/G</sup> round spermatids. Results shown in Fig. 4d demonstrate that H3K9 methylation levels were significantly increased in the promoter regions of both genes in *Jhdm2a*<sup>G/G</sup> spermatids relative to that in *Jhdm2a*<sup>+/+</sup> spermatids. However, the changes of H3K9 methylation levels were much less pronounced in a region ~1 kb upstream of the transcription start site (5') of the corresponding genes (Fig. 4d). Interestingly, a significant increase in the H3K9me3 level at both gene promoters is observed in the *Jhdm2a*<sup>G/G</sup> spermatids, despite previous demonstration that JHDM2A is incapable of removal of H3K9me3 *in vitro*<sup>5</sup>. It is possible that JHDM2A may associate with a co-factor that confers its substrate specificity when bound to the two promoters. Collectively, the above results support the notion that JHDM2A directly contributes to the expression of *Tnp1* and *Prm1* genes by binding to and removing the silencing H3K9 methyl mark at their promoters.

Through a loss-of-function approach, we demonstrated that *Jhdm2a* is indispensable for spermiogenesis. This function is consistent with its unique expression pattern in round and elongating spermatids. Our studies suggest that JHDM2A participates in spermiogenesis by regulating the expression of specific target genes such as *Tnp1* and *Prm1*, which are required for histone replacement during the final stages of sperm chromatin condensation and maturation (Supplementary Fig. 7). It is important to point out that, although residual amounts of the wild-type JHDM2A are still expressed in the *Jhdm2a*<sup>G/G</sup> mouse, and we cannot rule out the possibility that the phenotype might be the result of a dominant negative from the JHDM2A–β-Gal fusion protein, we observed the same phenotype in mice with a *Jhdm2a* allele that disrupts the JmjC domain, indicating that the *Jhdm2a*<sup>G/G</sup> allele phenocopies a true null (Y.O. and Y.Z., unpublished data). The fact that spermatogenesis is under the control of the hypothalamic–pituitary axis and is governed by hormones such as follicle-stimulating hormone and luteinizing hormone raises



**Figure 3 | Defective chromatin condensation in *Jhdm2a*<sup>G/G</sup> spermatids.**

**a**, Haematoxylin and eosin staining shows the histopathological features of testis from seven-month-old *Jhdm2a*<sup>+/+</sup> and *Jhdm2a*<sup>G/G</sup> mice. Roman numerals refer to the stages of mouse spermatogenesis. In *Jhdm2a*<sup>G/G</sup> mouse testis, spermatids with abnormal elongation were initially observed at stage XI (step 11). Insets represent higher-magnification images of elongated spermatids. **b**, Electron microscopy analysis of *Jhdm2a*<sup>+/+</sup> and *Jhdm2a*<sup>G/G</sup> spermatids. Chromocentre at step 2–3 of *Jhdm2a*<sup>G/G</sup> spermatids is indistinct. In steps 7–9, heterochromatin distribution loses polarity. In step 13, chromatin condensation is defective. **c**, Abnormal morphology of spermatozoa isolated from cauda epididymis of *Jhdm2a*<sup>G/G</sup> compared to spermatozoa from *Jhdm2a*<sup>+/+</sup>. *Jhdm2a*<sup>G/G</sup> spermatozoa exhibit round shaped heads and variable size. **d**, Sperm motility analysis of the indicated genotypes. **e**, Acridine orange staining of spermatozoa from *Jhdm2a*<sup>+/+</sup> and *Jhdm2a*<sup>G/G</sup> mice. Spermatozoa with complete chromatin condensation are only stained green, but are stained both green and red if their chromatin condensation is incomplete. Arrows indicate individual sperm. Blue staining represents Hoechst.

questions of whether the spermiogenesis defects observed in the *Jhdm2a* mutant mice are a secondary effect of hormonal or physiological pathway alterations. This is unlikely to be the case because no significant alternations in hormone levels are detected in the mutant mice (Supplementary Fig. 8). Consistently, no obvious morphological defects in Leydig or Sertoli cells are observed (Supplementary Fig. 3b). Indeed, we have not observed any additional phenotypes for the *Jhdm2a*<sup>G/G</sup> mice, supporting the contention that a main function of JHDM2A is in spermatogenesis. Although there are several knockout mouse models that exhibit male infertility phenotypes similar to those

observed in human syndromes such as azoospermia or globozoospermia<sup>21</sup>, most of the genes disrupted in mice are intact in human patients, raising the possibility that *Jhdm2a* might be a new candidate gene involved in these infertility syndromes.

## METHODS SUMMARY

**Generation of *Jhdm2a*<sup>G/G</sup> mice.** An ES clone with *Jhdm2a* gene-trap allele (YHA186) was purchased from BayGenomics (<http://baygenomics.ucsf.edu/>). The ES cells were injected into C57BL/6 blastocysts using standard procedures. After confirmation of germ-line transmission, the progeny was backcrossed to C57BL/6 once, and the *Jhdm2a*<sup>G/G</sup> line was obtained by crossing heterozygous (*Jhdm2a*<sup>+/-</sup>) mice.

**Histology, immunostaining and  $\beta$ -galactosidase staining.** For paraffin-embedded sections, dissected testes were fixed with either Bouin's fixative or 4% paraformaldehyde. For frozen sections, tissues were briefly fixed with 4% paraformaldehyde, dehydrated by sucrose, and embedded in optimal cutting temperature (OCT) compound (SAKURA). For non-fluorescent detection, an RTU vector staining kit was used (Vector laboratories). For  $\beta$ -galactosidase staining, tissues were fixed with 0.2% glutaraldehyde/2% formaldehyde buffer for 20 min. After rinsing with 0.1 M phosphate buffer (pH 7.3), they were stained with 1 mg ml<sup>-1</sup> 5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactoside (X-gal), 5 mM potassium ferricyanide and 5 mM potassium ferrocyanide in rinsing buffer at 37 °C for 4 h to overnight.

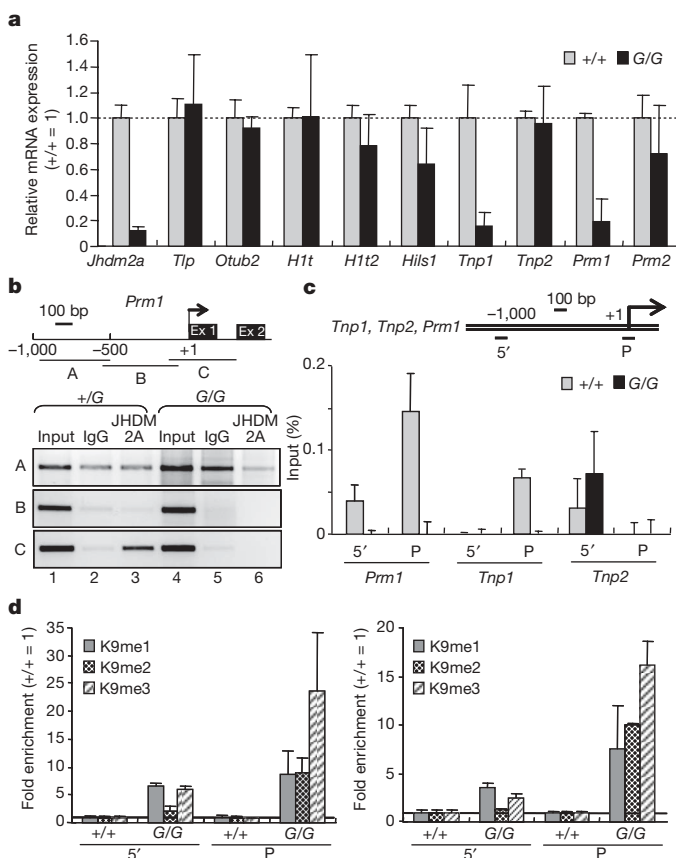
**RT-PCR and qRT-PCR.** Total RNA was purified from testes using a RNeasy kit (Qiagen). After DNase I treatment, first-strand DNA synthesis was performed using Imprim II (Promega). Ex Taq polymerase (TAKARA) and SYBR Green PCR master mix (Applied Biosystems) were used for PCR and qPCR, respectively. PCR conditions and primer sequences are listed in Supplementary Table 2.

**ChIP and ChIP/qPCR.** Spermatogenic cell fractionation was performed as described<sup>22</sup>. Cells recovered from each gradient fraction were partially stained by 4,6-diamidino-2-phenylindole (DAPI) to determine their purity. Selected fractions were pooled for specific developmental stages and their purity was further verified by RT-PCR using stage-specific marker genes (Supplementary Fig. 6). Cells ( $1 \times 10^6$  and  $5 \times 10^6$ ) were used for ChIP analysis with modification-specific anti-histone antibodies and anti-JHDM2A antibodies as described previously<sup>23</sup>. Ex Taq polymerase and SYBR Green PCR master mix were used for PCR and qPCR, respectively. PCR conditions and primer sequences are listed in Supplementary Table 2.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 13 July; accepted 5 September 2007.

Published online 17 October 2007.



**Figure 4 | JHDM2A positively regulates *Tnp1* and *Prm1* genes in round spermatids.** **a**, Quantitative RT-PCR analysis of the expression levels of selected genes in round spermatids from mice of indicated genotypes. *Otub2*, OTU domain, ubiquitin aldehyde binding 2. For *Jhdm2a*, primers 24F and 25R shown in Fig. 2a were used. Results were normalized to *Actb*, and the expression level in *Jhdm2a*<sup>+/+</sup> (+/+) is defined as 1. Data are presented as means  $\pm$  s.d. from three or four independent experiments. **b**, Top panel, diagram of mouse *Prm1* gene structure. A region 1 kb upstream of the transcription start site and first exon is covered by three primer sets (A, -980—-486; B, -551—-55; C, -124—+326). Bottom panel, ChIP analysis indicates that JHDM2A is recruited to a region that overlaps with the transcriptional start site of the *Prm1* gene (region 'C'). Input, 0.25%; IgG, rabbit immunoglobulin G; JHDM2A, anti-JHDM2A antibody. **c**, Top panel, diagram of mouse *Tnp1*, *Tnp2* and *Prm1* genes, and location of primers used for ChIP-qPCR. Bottom panel, ChIP-qPCR analysis examining recruitment of JHDM2A to promoter (P) and 1 kb upstream of transcriptional start site (5') of *Prm1*, *Tnp1* and *Tnp2* genes. The results were normalized to IgG control, and were shown as percentage of enrichment relative to input. Signals detected in the 5' of *Tnp2* are from non-specific PCR amplification. Data are presented as means  $\pm$  s.d. from three independent experiments. **d**, Comparison of H3K9 methylation levels of JHDM2A-binding sites of *Prm1* and *Tnp1* genes in *Jhdm2a*<sup>+/+</sup> (+/+) and *Jhdm2a*<sup>G/G</sup> (G/G) round spermatids by ChIP-qPCR. K9me1, K9me2 and K9me3 indicate anti-monomethyl-, dimethyl- and trimethyl H3K9 antibody, respectively. Results were normalized to histone H3, and are shown as fold of enrichment compared to that in the wild type (+/+). Data are presented as means  $\pm$  s.d. from three independent experiments.

- Klose, R. J. & Zhang, Y. Regulation of histone methylation by demethylimination and demethylation. *Nature Rev. Mol. Cell Biol.* **8**, 307–318 (2007).
- Shi, Y. & Whetstone, J. R. Dynamic regulation of histone lysine methylation by demethylases. *Mol. Cell* **25**, 1–14 (2007).
- Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941–953 (2004).
- Tsukada, Y. *et al.* Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**, 811–816 (2006).
- Yamane, K. *et al.* JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. *Cell* **125**, 483–495 (2006).
- Hoog, C., Schalling, M., Grunder-Brundell, E. & Daneholt, B. Analysis of a murine male germ cell-specific transcript that encodes a putative zinc finger protein. *Mol. Reprod. Dev.* **30**, 173–181 (1991).
- La Salle, S. & Trasler, J. M. Dynamic expression of DNMT3a and DNMT3b isoforms during male germ cell development in the mouse. *Dev. Biol.* **296**, 71–82 (2006).
- Kosower, N. S., Katayose, H. & Yanagimachi, R. Thiol-disulfide status and acridine orange fluorescence of mammalian sperm nuclei. *J. Androl.* **13**, 342–348 (1992).
- Traina, M. E. *et al.* Long-lasting effects of lindane on mouse spermatogenesis induced by *in utero* exposure. *Reprod. Toxicol.* **17**, 25–35 (2003).
- Peters, A. H. *et al.* Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**, 323–337 (2001).
- Payne, C. & Braun, R. E. Histone lysine trimethylation exhibits a distinct perinuclear distribution in Plzf-expressing spermatogonia. *Dev. Biol.* **293**, 461–472 (2006).
- Meistrich, M. L., Trostle-Weige, P. K., Lin, R., Bhatnagar, Y. M. & Allis, C. D. Highly acetylated H4 is associated with histone displacement in rat spermatids. *Mol. Reprod. Dev.* **31**, 170–181 (1992).
- Loukinov, D. I. *et al.* BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain

- with CTCF, the insulator protein involved in reading imprinting marks in the soma. *Proc. Natl Acad. Sci. USA* **99**, 6806–6811 (2002).
14. Seki, Y. *et al.* Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev. Biol.* **278**, 440–458 (2005).
  15. Sassone-Corsi, P. Unique chromatin remodeling and transcriptional regulation in spermatogenesis. *Science* **296**, 2176–2178 (2002).
  16. de Rooij, D. G. & de Boer, P. Specific arrests of spermatogenesis in genetically modified and mutant mice. *Cytogenet. Genome Res.* **103**, 267–276 (2003).
  17. Cho, C. *et al.* Haploinsufficiency of protamine-1 or -2 causes infertility in mice. *Nature Genet.* **28**, 82–86 (2001).
  18. Martianov, I. *et al.* Polar nuclear localization of H1T2, a histone H1 variant, required for spermatid elongation and DNA condensation during spermiogenesis. *Proc. Natl Acad. Sci. USA* **102**, 2808–2813 (2005).
  19. Martianov, I. *et al.* Late arrest of spermiogenesis and germ cell apoptosis in mice lacking the TBP-like *TLF/TRF2* gene. *Mol. Cell* **7**, 509–515 (2001).
  20. Zhao, M. *et al.* Transition nuclear proteins are required for normal chromatin condensation and functional sperm development. *Genesis* **38**, 200–213 (2004).
  21. Matzuk, M. M. & Lamb, D. J. Genetic dissection of mammalian fertility pathways. *Nature Cell Biol.* **4** (Suppl.), s41–s49 (2002).
  22. Pivot-Pajot, C. *et al.* Acetylation-dependent chromatin reorganization by BRDT, a testis-specific bromodomain-containing protein. *Mol. Cell. Biol.* **23**, 5354–5365 (2003).
  23. Okada, Y. *et al.* Leukaemic transformation by CALM-AF10 involves upregulation of Hoxa5 by hDOT1L. *Nature Cell Biol.* **8**, 1017–1024 (2006).
- Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).
- Acknowledgements** We thank T. Ward, V. Madden, R. Bagnell Jr and K. Moore for technical assistance, and K. Yamane for anti-JHDM2B antibody. We are grateful to R. Klose, E. Kallin and K. Gardner for reading of the manuscript. This work was supported by the NIH (Y.Z.), and in part by the Intramural Research Program of the NIH, NIEHS (Y.M.). Y.Z. is an Investigator of the Howard Hughes Medical Institute.
- Author Contributions** Y.O. and Y.Z. designed the experiments and prepared the manuscript. Y.O. performed the experiments. G.S., M.K.R. and Y.M. generated the chimaera mice from the BayGenomics ES clone.
- Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to Y.Z. (yi\_zhang@med.unc.edu).



## METHODS

**Generation of *Jhdm2a*<sup>G/G</sup> mice.** All mouse experiments were performed in accordance with institute guidelines covering the humane care and use of animals in research. Genotyping was performed by Southern blotting using 2 µg of tail DNA digested with either *EcoRV* or *SphI* as shown in Supplementary Fig. 2. The probe DNA was amplified by PCR.

**Antibodies.** Antibodies used for immunofluorescence, ChIP and western blotting are as follows: anti-JHDM2A<sup>3</sup>, anti-JHDM2B (rabbit polyclonal antibody raised against amino acids 165–465 of human JHDM2B), anti-H3K9me1 (Abcam, ab9045), anti-H3K9me2 (Upstate, 07-441), anti-H3K9me3 (Abcam, ab8898), anti-H3K4me3 (Abcam, ab8580), anti-acH4 (Upstate, 06-866), anti-β-Gal (MPBiologicals, 55976), anti-DNMT3B (a gift from E. Li), anti-γ-H2AX (Abcam, ab22551), anti-5' methylcytosine (Eutogentec, BI-MECY) and anti-RNA polymerase II (clone H14) (Covance, MMS-134R).

**Histology, immunostaining, β-galactosidase and acridine orange staining.** For β-galactosidase staining, tissues were fixed with 0.2% glutaraldehyde, 2% formaldehyde, 5 mM EGTA and 2 mM MgCl<sub>2</sub> in 0.1 M phosphate buffer (pH 7.3) for 20 min. After rinsing with 0.1% sodium deoxycholate, 0.2% NP-40 and 2 mM MgCl<sub>2</sub> in 0.1 M phosphate buffer (pH 7.3), they are stained with 1 mg ml<sup>-1</sup> X-gal, 5 mM potassium ferricyanide and 5 mM potassium ferrocyanide in rinsing buffer at 37 °C for 4 h to overnight (~12 h). Acridine orange staining was performed as previously described<sup>8</sup>.

**Electron microscopy.** Dissected testes were fixed with 2% paraformaldehyde and 2.5% glutaraldehyde in 0.15 M sodium phosphate buffer (pH 7.4) and transferred onto Epon. Ultrathin sections were stained with both uranyl acetate and lead citrate, and examined with an electron microscope (LEO EM910).

**ChIP and ChIP/qPCR.** Selected fractions were pooled for specific developmental stages and their purity was further verified by RT-PCR using stage-specific marker genes including *Sycp2* (pachytene spermatocytes), *Brd2* (round spermatids), *Prm2* (round and elongated spermatids), *Aqp8* (elongated spermatids) and *Actb* (housekeeping functions) (Supplementary Fig. 6).

**Flow cytometry and sperm-motility analysis.** Single-cell preparation and propidium-iodide staining were performed as previously described<sup>9</sup>. The motility of sperm isolated from cauda epididymis was measured by Mutant Mouse Regional Resource Centre at University of North Carolina using Computer Assistant Semen Analysis System (IVOS, Hamilton-Thorne Research).

**Hormone measurement.** Follicle-stimulating hormone, leutinizing hormone, oestradiol, testosterone and androstenedione levels in mouse serum were examined by the Ligand Core Laboratory at University of Virginia.

## LETTERS

# Probing the chemistry of thioredoxin catalysis with force

Arun P. Wiita<sup>1,2</sup>, Raul Perez-Jimenez<sup>1</sup>, Kirstin A. Walther<sup>1,3</sup>, Frauke Gräter<sup>4</sup>, B. J. Berne<sup>4</sup>, Arne Holmgren<sup>5</sup>, Jose M. Sanchez-Ruiz<sup>6</sup> & Julio M. Fernandez<sup>1</sup>

Thioredoxins are enzymes that catalyse disulphide bond reduction in all living organisms<sup>1</sup>. Although catalysis is thought to proceed through a substitution nucleophilic bimolecular ( $S_N2$ ) reaction<sup>1,2</sup>, the role of the enzyme in modulating this chemical reaction is unknown. Here, using single-molecule force-clamp spectroscopy<sup>3,4</sup>, we investigate the catalytic mechanism of *Escherichia coli* thioredoxin (Trx). We applied mechanical force in the range of 25–600 pN to a disulphide bond substrate and monitored the reduction of these bonds by individual enzymes. We detected two alternative forms of the catalytic reaction, the first requiring a reorientation of the substrate disulphide bond, causing a shortening of the substrate polypeptide by  $0.79 \pm 0.09 \text{ \AA}$  ( $\pm$  s.e.m.), and the second elongating the substrate disulphide bond by  $0.17 \pm 0.02 \text{ \AA}$  ( $\pm$  s.e.m.). These results support the view that the Trx active site regulates the geometry of the participating sulphur atoms with sub-ångström precision to achieve efficient catalysis. Our results indicate that substrate conformational changes may be important in the regulation of Trx activity under conditions of oxidative stress and mechanical injury, such as those experienced in cardiovascular disease<sup>5,6</sup>. Furthermore, single-molecule atomic force microscopy techniques, as shown here, can probe dynamic rearrangements within an enzyme's active site during catalysis that cannot be resolved with any other current structural biological technique.

One of the principal challenges of understanding enzyme catalysis, a central problem in biology, is resolving the dynamics of enzyme–substrate interactions with sub-ångström resolution—the length scale at which chemistry occurs<sup>7</sup>. Although nuclear magnetic resonance (NMR) and X-ray crystallography determinations of protein structures can reach down to the sub-ångström level, they cannot yet provide dynamic information about enzyme catalysis at this length scale<sup>8</sup>. Here we demonstrate the ability of single-molecule techniques in probing the dynamics of enzyme catalysis at the sub-ångström scale.

We used a polypeptide made of eight repeats of the I27 domain of human cardiac titin<sup>9</sup> with engineered cysteines, (I27<sub>SS</sub>)<sub>8</sub>, as a substrate protein to monitor the Trx-catalysed reduction of individual disulphide bonds (SS) placed under a stretching force. In our experiments, we used atomic force microscopy in force-clamp mode<sup>3</sup> to extend single (I27<sub>SS</sub>)<sub>8</sub> polypeptides (Fig. 1a, far left). The constant applied force caused individual domains to unfold, resulting in a stepwise increase in the length of the molecule after each unfolding event. This is illustrated in Fig. 1b, in which a single (I27<sub>SS</sub>)<sub>8</sub> polypeptide was first mechanically unfolded at 165 pN for 400 ms. A series of 10.8-nm steps was rapidly observed (Fig. 1b, inset); each step corresponds to the partial unfolding of a single I27<sub>SS</sub> domain up to the disulphide bond (red, Fig. 1a). The disulphide bond is buried in

the folded protein<sup>4</sup> and is exposed to the bathing solution only after partial unfolding. The unfolding force pulse was followed by a test pulse (100 pN in this case). No further steps were observed during the test pulse because the disulphide bond could not be broken by the applied force alone<sup>10</sup> in the absence of Trx. After unfolding, the stretching force was applied directly to the disulphide bond and, if Trx is present in solution, the bond can be chemically reduced by the enzyme. Such a result is shown in Fig. 1c, with a similar experiment in the presence of 8  $\mu\text{M}$  Trx. Now, during the test pulse, seven steps of  $\sim 13.2 \text{ nm}$  were observed as individual disulphide bonds were reduced by single Trx enzymes, allowing for the immediate extension of the residues previously trapped behind the disulphide bond (blue, Fig. 1a). The size of the increases in step length observed during these force-clamp experiments corresponds to the number of amino acids released, serving as a precise fingerprint to identify the reduction events<sup>4,11</sup>.

We used an ensemble of single-molecule recordings to measure the kinetics of disulphide bond reduction by Trx. At each force and Trx concentration, we averaged (Supplementary Fig. 1) 10–30 test-pulse recordings of the type shown in Fig. 1c. Averaged traces at various forces are shown in Fig. 2a. The averaged traces were fitted by a single exponential with a time constant  $\tau$  (see Supplementary Figs 2 and 3). We define the observed rate constant of disulphide reduction as  $r = 1/\tau$ .

Figure 2b shows a plot of  $r$  as a function of the applied (test pulse) force. The figure shows that the rate of reduction decreases fourfold between 25 and 250 pN, and then increases approximately threefold when the force is increased up to 600 pN, demonstrating a biphasic force dependency. This result is in contrast with the uniform acceleration of dithiothreitol (DTT) reduction rate with increasing force<sup>4</sup>, underlining a much more complex chemical reaction catalysed by Trx. Furthermore, the rate of reduction becomes saturated as the concentration of Trx is increased (Fig. 2c).

To explain our data, we tested different kinetic models of force-dependent Trx catalysis (Supplementary Figs 4–6, and Supplementary Tables 2 and 3). We found that the model that could best describe our data incorporates an intermediate state as well as two different force-dependent rate constants (Fig. 2d). Path I (red, Fig. 2d) is similar to a Michaelis–Menten mechanism, with a catalytic step inhibited by force. Path II (blue, Fig. 2d) is governed solely by the rate constant  $k_{02}$  (where subscripts refer to steps in Fig. 2d), which is accelerated by force. Our model can be globally fitted to the data of Fig. 2b and c (solid lines), obtaining values for the model parameters (Supplementary Table 1). The goodness of fit for this model was measured using statistical methods<sup>12</sup> (see Methods and Supplementary Table 3;  $\chi^2_\nu = 0.835$ , and  $\nu = 26$ , six free parameters;  $P(\chi^2_\nu) = 0.705$ ). An extrapolation to zero force predicts a second-order rate constant for Trx reduction of  $2.2 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ . This is  $\sim 30,000$  times faster than that

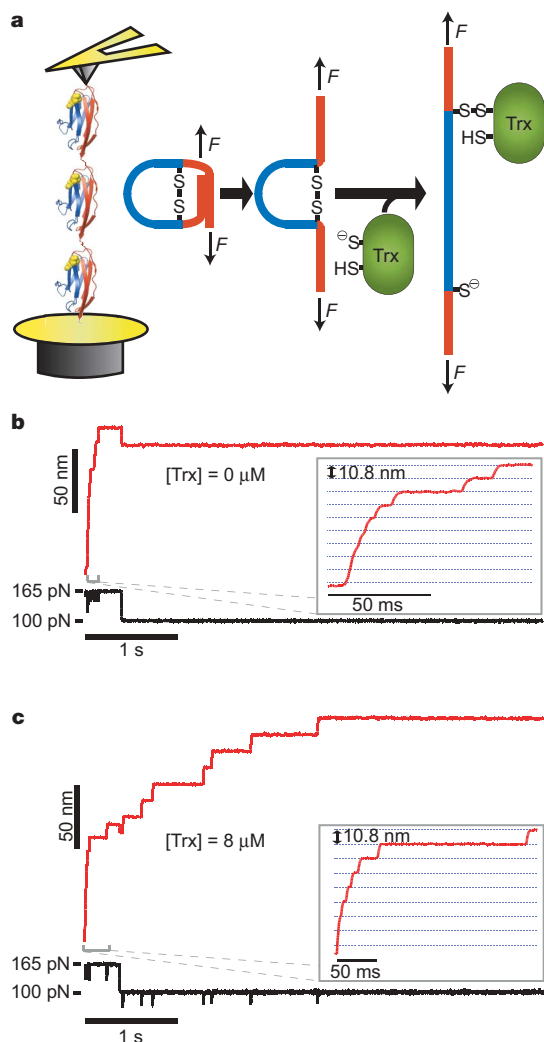
<sup>1</sup>Department of Biological Sciences, <sup>2</sup>Graduate Program in Neurobiology and Behavior, <sup>3</sup>Department of Physics, <sup>4</sup>Department of Chemistry, Columbia University, New York, New York 10027, USA. <sup>5</sup>Medical Nobel Institute for Biochemistry, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, SE-171 77, Stockholm, Sweden. <sup>6</sup>Facultad de Ciencias, Departamento de Química Física, Universidad de Granada, 18071, Granada, Spain.

found for I27<sub>SS</sub> disulphide reduction by DTT ( $6.5 \text{ M}^{-1} \text{ s}^{-1}$ , ref. 4). This result is consistent with bulk biochemical experiments, in which Trx has been found to reduce insulin disulphide bonds  $\sim 20,000$  times faster than DTT ( $1 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$  for Trx versus  $5 \text{ M}^{-1} \text{ s}^{-1}$  for DTT at pH 7 (ref. 13)).

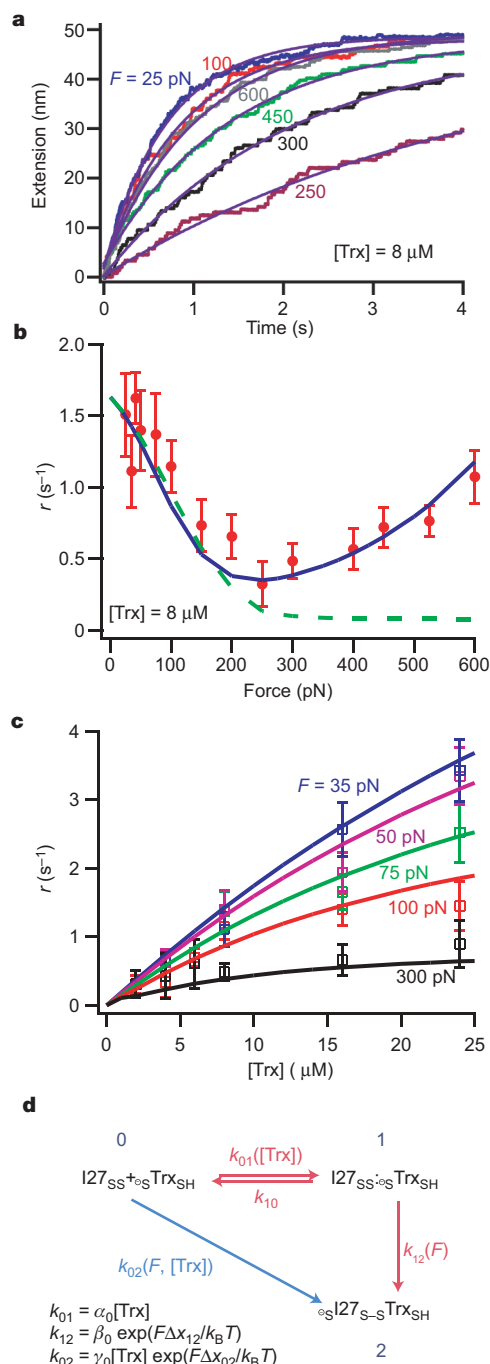
The experimental data shown in Fig. 2b suggest that there are two separate pathways for disulphide bond reduction by Trx. Further support for this hypothesis was gained by probing the force-dependent reduction kinetics of an active site mutant, Trx(P34H) (Fig. 3). In our single-molecule experiments, the extrapolated zero-force rate of reduction for Trx(P34H) is less than one-half of that for the wild-type enzyme ( $8.8 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$  versus  $2.2 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ ), showing a similar relationship to bulk biochemical experiments ( $3 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$  for Trx(P34H) versus  $2 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$  for wild-type Trx at pH 8 and  $15^\circ \text{C}$  (ref. 14)). In Trx(P34H), the rate of Trx binding to the substrate ( $k_{01}$ ) decreased significantly, whereas the other kinetic parameters remain mostly unchanged (Supplementary Table 1). By fitting

this data with two alternate kinetic models, we found that the Trx(P34H) mutant supports the view that Trx has two distinct forms of catalysis, without a common intermediate (see Supplementary Figs 5–7 and Supplementary Table 3).

In the kinetic model shown in Fig. 2d, the catalytic rate constants are described by a straightforward Arrhenius term. For example,



**Figure 1 | Identification of single Trx catalytic events.** **a**, Single (I27<sub>SS</sub>)<sub>8</sub> molecules were stretched using an atomic force microscope in force-clamp mode (left). After unfolding of the red residues, the disulphide bond is exposed to the solution. On disulphide reduction by Trx, the blue residues previously trapped behind the disulphide bond are immediately extended. **b**, A single (I27<sub>SS</sub>)<sub>8</sub> molecule is stretched in the absence of Trx. At least six individual domains are unfolded up to the disulphide bond during the unfolding pulse to 165 pN (inset). No further steps are noted during the test pulse to 100 pN. **c**, In the presence of  $8 \mu\text{M}$  Trx, seven steps of  $\sim 13.2 \text{ nm}$  are observed during the test pulse, corresponding to the extension of the trapped residues in each module after the reduction of individual disulphide bonds by single Trx enzymes.



**Figure 2 | Force-dependent Trx catalysis.** **a**, Multiple single-molecule recordings of the test pulse only ( $n = 10\text{--}30$ ) were averaged to monitor the kinetics of disulphide bond reduction under force  $F$ . A single exponential is fitted to each averaged trace (smooth line), and the rate constant of reduction  $r = 1/\tau$ . **b**,  $r$  as a function of force at  $[\text{Trx}] = 8 \mu\text{M}$ . **c**,  $r$  as a function of  $[\text{Trx}]$  at various forces. Error bars in **b** and **c** represent the s.e.m. obtained from bootstrapping (see Methods). Solid lines in **b** and **c** are fits using the kinetic model shown in **d**. This model describes two modes of Trx catalysis (path I in red, path II in blue), where the catalytic rate constants are exponentially dependent on the applied force. The dashed green line in **b** represents model fits in the absence of path II, with  $k_{02} = 0$ .



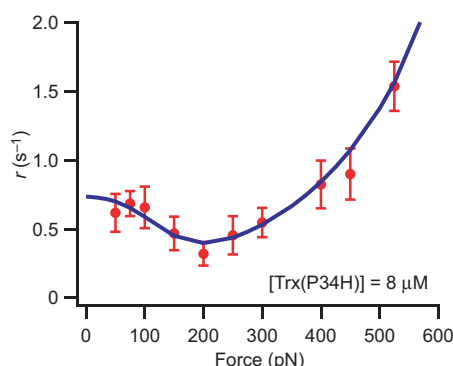
$k_{12} = \beta_0 \exp(F\Delta x_{12}/k_B T)$ , where  $\beta_0$  is the rate constant at zero force,  $k_B$  is Boltzmann's constant,  $T$  is the temperature and  $\Delta x_{12}$  is the distance to the transition state along the length coordinate<sup>15</sup>. Fits of the kinetic model (Fig. 2d) to the data of Fig. 2b and c gave values of  $\Delta x_{12} = -0.79 \pm 0.09$  Å for the catalytic step of path I and  $\Delta x_{02} = 0.17 \pm 0.02$  Å for the catalytic step of path II ( $\pm$  s.e.m. obtained by downhill simplex procedure for model fitting, see Methods). Similar parameters were also found for the Trx(P34H) mutant (Supplementary Table 1). Thus, the two catalytic pathways are very different: the transition state of reduction by way of path I requires a shortening of the substrate polypeptide by  $\sim 0.8$  Å, whereas path II requires an elongation by  $\sim 0.2$  Å.

Our experiments show that sub-ångström-level distortions of the substrate disulphide bond take place dynamically during Trx catalysis. A glimpse of the transition state for Trx catalysis can be obtained from the NMR structure of human TRX (also known as TXN) a homologue of the *E. coli* enzyme<sup>16,17</sup>, in a complex with a substrate peptide from the signalling protein NF- $\kappa$ B (Fig. 4a, PDB accession number 1MDI). In this structure (ref. 18), as well as in the structure of human TRX bound to REF-1 (also known as APEX1) (ref. 19), a peptide-binding groove is identified on the surface of TRX in the vicinity of the catalytic Cys 32. The sulphur atom in Cys 32 (sulphur atom A) of the active site of TRX forms a disulphide bond with the sulphur atom of the NF- $\kappa$ B peptide (sulphur atom B).

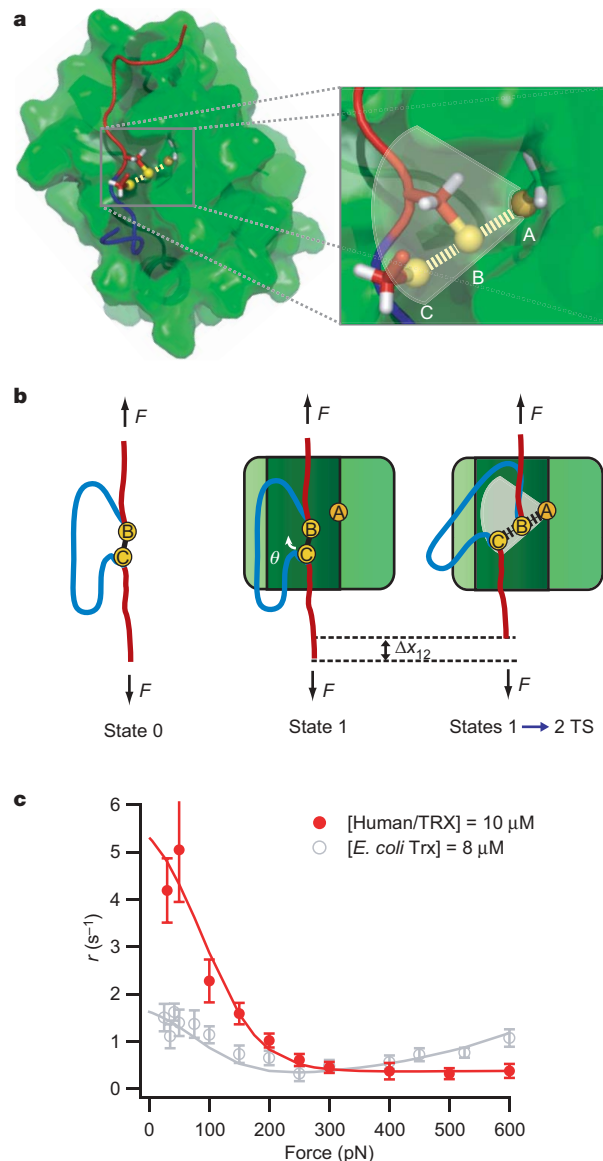
We used the orientation of the disulphide bond within the Trx active site in an attempt to predict the structure of the catalytic transition state in our experiments. It is known that disulphide bond reduction proceeds by means of an  $S_N2$  mechanism. This reaction is highly directional, proceeding via a transition state in which the three involved sulphur atoms form an  $\sim 180^\circ$  angle<sup>20–22</sup>. Thus, the relative positions of these sulphur atoms must be important for efficient Trx catalysis. We found that the disulphide bond in 1MDI forms an angle of  $\sim 70^\circ$  with respect to the axis of the peptide-binding groove. Assuming that this orientation applies to the  $S_N2$  reaction that reduces the I27<sub>SS</sub> bond of our experiments, and that the stretched polypeptide is bound to the groove, it is apparent that the target disulphide bond must rotate with respect to the pulling axis to acquire the correct  $S_N2$  geometry (Fig. 4b). Given that the disulphide bond in the stretched polypeptide is aligned within  $\sim 20^\circ$  of the pulling force (Supplementary Fig. 8), a further rotation by an angle  $\theta = 50^\circ$  would be required for catalysis (Fig. 4b), causing a contraction of the target polypeptide by  $\sim 1.2$  Å, close to the measured value of  $\Delta x_{12} \approx -0.8$  Å. However, molecular dynamics simulations have previously identified multiple conformations of the catalytic thiol in glutaredoxin, a member of the thioredoxin superfamily<sup>23</sup>.

Similarly, we have performed molecular dynamics simulations of the 1MDI structure to examine the conformational diversity of the

NF- $\kappa$ B to Cys 32 disulphide bond. Our simulations show that the disulphide bond samples a range of conformations with  $\theta = 50^\circ - 80^\circ$  in either the clockwise or the counterclockwise direction (shaded area in the inset of Fig. 4a, and Supplementary Figs 9 and 10). We then combined the results of these molecular dynamics simulations with a theoretical model that treats the substrate backbone as a freely jointed chain<sup>24</sup>. This model predicts the likelihood of the substrate disulphide achieving the correct geometry for the reaction transition state under a pulling force (see Supplementary Information and Supplementary Fig. 10). We found that, in the cases of NF- $\kappa$ B, REF-1 and the apo TRX, an average bond rotation on the



**Figure 3 | Trx(P34H).** Plot of  $r$  as a function of force at  $[\text{Trx(P34H)}] = 8 \mu\text{M}$ . Error bars represent the s.e.m. obtained from bootstrapping (see Methods). The solid line is the best fit obtained with the three-state model of Fig. 2d. This point mutation significantly decreases parameter  $k_{01}$ —the rate of enzyme binding to the substrate—whereas other parameters in the kinetic model remain similar to wild-type Trx (see Supplementary Table 1).



**Figure 4 | Structural model for force-dependent Trx catalysis.** **a**, TRX (peptide-binding groove in dark green) bound to an NF- $\kappa$ B peptide. The inset (yellow spheres are sulphur atoms A, B and C) shows the relative position of the disulphide bond between TRX Cys 32 (sulphur atom A) and the NF- $\kappa$ B cysteine (sulphur atom B). The third sulphur atom (sulphur atom C) belonging to the leaving cysteine was placed  $180^\circ$  from the disulphide bond, as required by the  $S_N2$  chemical reaction. **b**, Cartoon representation of the reduction by Trx of a disulphide bond in a stretched polypeptide. On binding, the substrate disulphide bond (between sulphur atoms B and C) has to rotate by an angle  $\theta$  to acquire the correct  $S_N2$  geometry at the transition state (TS) of the reaction, causing a contraction of the substrate polypeptide by an amount  $\Delta x_{12}$ . This rotation is opposed by the pulling force. **c**, Force-dependent reduction by human TRX compared to *E. coli* Trx. Error bars represent the s.e.m. obtained from bootstrapping (see Methods).

sub-ångström scale (resulting in  $\Delta x_{12}$  values of  $-0.77$ ,  $-0.45$  and  $-0.19$  Å, respectively) must take place to allow  $S_N2$  chemistry in the TRX active site (Supplementary Table 4).

To probe this model of catalysis, which is based on the structure of human TRX complexes, we also tested the force-dependent mechanism of disulphide bond reduction by human TRX (Fig. 4c). At low forces, it is clear that human TRX catalyses disulphide bond reduction in I27<sub>SS</sub> much more rapidly than Trx from *E. coli*. However, at high forces it appears that path II is quite diminished in the human TRX variant. Thus, the data for human TRX resemble a simple Michaelis–Menten model (dashed green line in Fig. 2b), indicating that the two thioredoxin variants differ in their catalytic mechanisms at high force. Our three-state kinetic model also describes the human TRX data well with a fixed  $\Delta x_{12} = -0.79$  Å (Fig. 4c and Supplementary Tables 1 and 3). Thus, it is clear that the mechanism that governs the force-dependence of path I is conserved between these homologues, and the results for human TRX at low forces can also be explained by our structural model.

The origin of the  $\Delta x_{02} \sim 0.2$  Å elongation at the transition state of catalysis for *E. coli* Trx, measured from the force-dependency of path II, is less clear. However, as demonstrated in the theoretical calculations of thiol/disulphide exchange in ref. 22, other reaction geometries are possible, even if they are typically unfavourable energetically. Thus,  $\Delta x_{02}$  may correspond to the lengthening of the I27<sub>SS</sub> disulphide bond at a transition state<sup>4</sup> other than the standard  $S_N2$  form.

Our results show that a mechanical force can alter the chemistry of the catalytic site in thioredoxin significantly. This is a novel concept in biology, that mechanical stresses applied to tissues may completely change the enzymatic chemistry from that observed in solution biochemistry. These effects may be particularly significant in tissues exposed to pathological force levels such as those that occur during mechanical injury. For example, it is well known that the increased mechanical stress during hypertension triggers an oxidative stress response in vascular endothelium and smooth muscle<sup>5</sup> that is compensated by an increase in the activity of thioredoxin<sup>6,25</sup>. In this context, we predict that the increased mechanical forces applied to target disulphide bonds would inhibit the activity of thioredoxin, diminishing the effectiveness of the antioxidant properties of the enzyme. The capability of single-molecule atomic force microscopy techniques directly to probe the dynamic sub-ångström molecular rearrangements during catalysis may prove to be an important tool in understanding the fundamental mechanisms underlying enzymatic chemistry.

## METHODS SUMMARY

The expression and purification of I27<sub>SS</sub>, wild-type Trx and Trx(P34H) are described in the Methods. Our custom-built atomic force microscope controlled by an analogue proportional-integral-derivative (PID) feedback system has been described previously<sup>3</sup>. The buffer used in the experiments contained 10 mM HEPES, 150 mM NaCl, 1 mM EDTA, 2 mM NADPH, 50 nM thioredoxin reductase (from *E. coli* for Trx and from rat liver for TRX) and the indicated concentration of Trx or TRX, and was controlled to pH 7.2. Single (I27<sub>SS</sub>)<sub>8</sub> protein molecules were stretched by first pressing the cantilever on the coverslide at a constant force of 800 pN for 3 s, then retracting to a constant force of 165 pN for 400 ms during the unfolding pulse. The indicated test-pulse force was applied for  $\sim 5$  s. For further details, see Methods. All data were obtained and analysed using custom software written for use in Igor 5.0 (Wavemetrics). We summed and normalized the test-pulse portions of numerous ( $n = 10$ –30) recordings that contained only disulphide reduction events and no unsequestered unfolding events to obtain the experimental value  $r$ . The differential rate equations were solved using matrix analysis methods, and error analysis was performed using the nonparametric bootstrap method in combination with the downhill simplex method (see Methods for details). All error bars shown represent standard error. See Methods for description of molecular dynamics simulations, and Supplementary Information for theoretical modelling of the transition state for Trx catalysis.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 21 May; accepted 7 September 2007.**

- Holmgren, A. Thioredoxin. *Annu. Rev. Biochem.* **54**, 237–271 (1985).
- Holmgren, A. Thioredoxin structure and mechanism: conformational changes on oxidation of the active-site sulfhydryls to a disulfide. *Structure* **3**, 239–243 (1995).
- Schlierf, M., Li, H. & Fernandez, J. M. The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques. *Proc. Natl Acad. Sci. USA* **101**, 7299–7304 (2004).
- Wiita, A. P., Ainarapu, S. R. K., Huang, H. H. & Fernandez, J. M. Force-dependent chemical kinetics of disulfide bond reduction observed with single-molecule techniques. *Proc. Natl Acad. Sci. USA* **103**, 7222–7227 (2006).
- Paravicini, T. M. & Touyz, R. M. Redox signaling in hypertension. *Cardiovasc. Res.* **71**, 247–258 (2006).
- World, C. J., Yamawaki, H. & Berk, B. C. Thioredoxin in the cardiovascular system. *J. Mol. Med.* **84**, 997–1003 (2006).
- Kraut, D. A., Carroll, K. S. & Herschlag, D. Challenges in enzyme mechanism and energetics. *Annu. Rev. Biochem.* **72**, 517–571 (2003).
- Hammes-Schiffer, S. & Benkovic, S. J. Relating protein motion to catalysis. *Annu. Rev. Biochem.* **75**, 519–541 (2006).
- Carrión-Vázquez, M. et al. Mechanical and chemical unfolding of a single protein: a comparison. *Proc. Natl Acad. Sci. USA* **96**, 3694–3699 (1999).
- Grandbois, M., Beyer, M., Rief, M., Clausen-Schaumann, H. & Gaub, H. E. How strong is a covalent bond? *Science* **283**, 1727–1730 (1999).
- Ainarapu, S. R. et al. Contour length and refolding rate of a small protein controlled by engineered disulfide bonds. *Biophys. J.* **92**, 225–233 (2007).
- Abbondanzieri, E. A., Greenleaf, W. J., Shaevitz, J. W., Landick, R. & Block, S. M. Direct observation of base-pair stepping by RNA polymerase. *Nature* **438**, 460–465 (2005).
- Holmgren, A. Reduction of disulfides by thioredoxin. Exceptional reactivity of insulin and suggested functions of thioredoxin in mechanism of hormone action. *J. Biol. Chem.* **254**, 9113–9119 (1979).
- Krause, G., Lundström, J., Barea, J. L., Pueyo de la Cuesta, C. & Holmgren, A. Mimicking the active site of protein disulfide-isomerase by substitution of proline 34 in *Escherichia coli* thioredoxin. *J. Biol. Chem.* **266**, 9494–9500 (1991).
- Bell, G. I. Models for the specific adhesion of cells to cells. *Science* **200**, 618–627 (1978).
- Qin, J., Clore, G. M. & Gronenborn, A. M. The high-resolution three-dimensional solution structures of the oxidized and reduced states of human thioredoxin. *Structure* **2**, 503–522 (1994).
- Eklund, H., Gleason, F. K. & Holmgren, A. Structural and functional relations among thioredoxins of different species. *Proteins* **11**, 13–28 (1991).
- Qin, J., Clore, G. M., Kennedy, W. P., Huth, J. R. & Gronenborn, A. M. Solution structure of human thioredoxin in a mixed disulfide intermediate complex with its target peptide from the transcription factor NFκB. *Structure* **3**, 289–297 (1995).
- Qin, J., Clore, G. M., Kennedy, W. P., Kuszewski, J. & Gronenborn, A. M. The solution structure of human thioredoxin complexed with its target from Ref-1 reveals peptide chain reversal. *Structure* **4**, 613–620 (1996).
- Rosenfield, R. E., Parthasarathy, R. & Dunitz, J. D. Directional preferences of nonbonded atomic contacts with divalent sulfur. 1. Electrophiles and nucleophiles. *J. Am. Chem. Soc.* **99**, 4860–4862 (1977).
- Pappas, J. A. Theoretical studies of reactions of sulfur–sulfur bond. 1. General heterolytic mechanisms. *J. Am. Chem. Soc.* **99**, 2926–2930 (1977).
- Fernandes, P. A. & Ramos, M. J. Theoretical insights into the mechanism for thiol/disulfide exchange. *Chem. Eur. J.* **10**, 257–266 (2004).
- Foloppe, N. & Nilsson, L. The glutaredoxin -C-P-Y-C- motif: influence of peripheral residues. *Structure* **12**, 289–300 (2004).
- Grosberg, A. Y. & Khokhlov, A. R. *Statistical Physics of Macromolecules* (AIP, New York, 1994).
- Tao, L. et al. Cardioprotective effects of thioredoxin in myocardial ischemia and reperfusion: role of S-nitrosation. *Proc. Natl Acad. Sci. USA* **101**, 11471–11476 (2004).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank D. Rodríguez-Larrea for thioredoxin purification and S. Posy for assistance with structural modelling. This work was supported by NIH grants to J.M.F., an NIH grant to B.J.B., a grant from the Swedish Society for Medical Research to A.H., and a grant from the Spanish Ministry of Science and Education to J.M.S.-R. F.G. is supported by an ISE Columbia University grant to J.M.F. and B.J.B. A.P.W. is supported by an NIH Medical Scientist Training Program grant to Columbia University.

**Author Contributions** A.P.W., R.P.-J. and J.M.F. designed the experiments. A.P.W. and R.P.-J. performed the experiments and analysed the data. K.A.W. designed the kinetic model and performed error analysis. F.G. and B.J.B. performed molecular dynamics simulations. A.H. provided TRX. J.M.S.-R. provided Trx and Trx(P34H). A.P.W., F.G., K.A.W., R.P.-J. and J.M.F. wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to J.M.F. ([jfernandez@columbia.edu](mailto:jfernandez@columbia.edu)).

## METHODS

**Protein engineering, expression and purification.** The expression and purification of (I27<sub>SS</sub>)<sub>8</sub> has been described previously<sup>4</sup>. In brief, we used the QuikChange site-directed mutagenesis method (Stratagene) to introduce Gly 32 Cys and Ala 75 Cys mutations into the I27 module from human cardiac titin. We used multiple rounds of successive cloning<sup>9</sup> to create an amino-carboxy linked, eight-domain polypeptide gene, (I27<sub>G32C-A75C</sub>)<sub>8</sub>. In this work, we call this construct (I27<sub>SS</sub>)<sub>8</sub>. This gene was encoded in vector pQE30 and expressed in *E. coli* strain BL21(DE3). Pelleted cells were lysed by sonication, and the His 6-tagged protein was first purified using an immobilized Talon-Co<sup>2+</sup> column (Clontech) and then further purified by gel filtration on a Superdex 200 column (GE Healthcare). The purified protein was verified by SDS-PAGE and stored at 4 °C in a buffer of 10 mM HEPES, 150 mM NaCl, 1 mM EDTA and 0.02% NaN<sub>3</sub> (w/v), pH 7.2.

Both wild-type Trx and Trx(P34H) were expressed and purified by the same method described previously<sup>26</sup>. Briefly, the *E. coli* Trx gene encoded in plasmid pTK100, was expressed in *E. coli* strain JF521. Cell pellets were lysed using a French press and stirred with streptomycin sulphate (10% w/v) at 4 °C for 16 h. The filtered supernatant was then loaded onto a 2-l Sephacryl S-100 High Resolution (GE Healthcare) gel filtration column. Trx fractions were pooled and applied to a 250-ml Fractogel EMD DEAE(M) (Merck) ion exchange column equilibrated in a buffer containing 1 mM EDTA and 30 mM TRIZMA, pH 8.3. The protein was eluted by a linear gradient between 0 and 0.5 M NaCl. The proteins were pure, as measured by SDS-PAGE gel densitometry. The molecular weight of pure proteins was confirmed by mass spectrometry. Trx fractions were dialysed into a buffer of 10 mM HEPES, 150 mM NaCl and 1 mM EDTA, pH 7.2. Trx concentration was determined spectrophotometrically at 280 nm using a molar absorption coefficient  $\epsilon_{280}$  of 13,700 M<sup>-1</sup> cm<sup>-1</sup> (ref. 27). The bulk activity of Trx and Trx(P34H) was confirmed by monitoring spectrophotometrically at 412 nm the reduction of 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB, Sigma) as described<sup>28</sup>.

TRX was purified as previously described<sup>29</sup>. Briefly, the pACA/TRX plasmid was expressed in BL21(DE3) cells. Cell pellets were lysed using a French press and stirred with 7% w/v streptomycin sulphate. Protein was then precipitated by adding ammonium sulphate to 85% saturation. The crude extracts were applied to a DEAE 52 column equilibrated with 50 mM Tris-HCl, pH 7.5, 1 mM EDTA and 0.1 mM DTT. Protein was eluted with an NaCl gradient, pooled and concentrated, and then applied to a Sephadex G-50 column equilibrated with 50 mM Tris-HCl, pH 7.5, 1 mM EDTA and 0.1 mM DTT. Fractions were pooled, concentrated and further purified using *E. coli* Trx antibody affinity chromatography. Protein concentration was determined spectrophotometrically at 280 nm using an  $\epsilon_{280}$  of 8,050 M<sup>-1</sup> cm<sup>-1</sup>.

**Single-molecule force-clamp spectroscopy.** Our custom-built atomic force microscope has been described previously<sup>30</sup>. Typical resolution in extension was ~0.5 nm and typical analogue feedback lag in the force-clamp following unfolding was ~5 ms. The spring constant of silicon nitride cantilevers (Veeco), typically ~20 pN nm<sup>-1</sup>, was calibrated as described previously<sup>31</sup>. The buffer used for all experiments contained 10 mM HEPES, 150 mM NaCl, 1 mM EDTA and 2 mM NADPH, and was controlled to pH 7.2. Before beginning the experiment, thioredoxin reductase (Sigma; from *E. coli* for *E. coli* Trx experiments, or from rat liver for human TRX experiments) was added to the experimental buffer to a final concentration of 50 nM. Thioredoxin was then added to the experimental buffer to the indicated concentration. An excess of NADPH and a catalytic amount of thioredoxin reductase are both necessary to maintain ~98% of Trx in the active, reduced form during the experiment<sup>1</sup>. In the Trx system, reducing equivalents are donated from NADPH to the FAD domain of thioredoxin reductase, and these electrons subsequently reduce a catalytic disulphide bond in thioredoxin reductase. Reduced *E. coli* thioredoxin reductase is very specific for reducing the disulphide bond in oxidized Trx<sup>1</sup> and does not non-specifically reduce other disulphides. This is demonstrated by recordings shown in Fig. 1b and Supplementary Fig. 1b; when Trx was not included in the solution, no disulphide reduction in I27<sub>SS</sub> was observed even in the presence of thioredoxin reductase and NADPH.

In the experiment, ~5  $\mu$ l (I27<sub>SS</sub>)<sub>8</sub> solution was added to a ~100  $\mu$ l droplet of Trx-containing experimental buffer deposited on a substrate coverslide. Single (I27<sub>SS</sub>)<sub>8</sub> protein molecules were stretched by first pressing the cantilever on the coverslide at a constant force of 800 pN for 3 s, then retracting to a constant force of 165 pN for 400 ms during the unfolding pulse. The indicated test-pulse force was applied for ~5 s. In our experiments we did not control the precise point of attachment between the (I27<sub>SS</sub>)<sub>8</sub> molecule and the cantilever; thus, varying numbers of disulphide reduction events may be observed in a given single-molecule recording.

**Data analysis.** All data were recorded and analysed using custom software written in Igor Pro 5.0 (Wavemetrics). We analysed only recordings that exhibited disulphide reduction events of the expected step size in the test pulse. (For a discussion of expected disulphide reduction step size as a function of force, see the Supporting Online Material of ref. 4.) We summated and normalized the test-pulse portions of numerous ( $n = 10$ –30) recordings that contained only disulphide reduction events and no unsequestered unfolding events. We fitted these averaged traces with a single exponential to obtain the observed rate constant of reduction,  $r$ . This type of summation procedure is standard in the ion channel literature and has been used in many contexts to obtain macroscopic kinetics from single-molecule recordings<sup>32,33</sup>. We assume that disulphide reduction in (I27<sub>SS</sub>)<sub>8</sub> is markovian (that is, that each reduction event is independent of all others); thus, averaging traces with different numbers of reduction steps will result in invariant exponential kinetics<sup>4</sup>. To estimate the error on our experimentally obtained rate constant, we carried out the nonparametric bootstrap method<sup>34</sup>. At a given value of force and [Trx],  $n$  staircases were randomly drawn with replacement from our original data set. These were summed and fitted to obtain a rate constant. This procedure was repeated 1,000 times for each data set, resulting in a distribution that provided the standard error of the mean for the reduction rate constant, shown as the error bars in Figs 2b and c, 3 and 4c.

**Kinetic model.** In the kinetic model shown in Fig. 2d (and Supplementary Fig. 4a), three states are used to describe our experimental system. The rate equations for the concentrations of states 0, 1 and 2 as a function of time  $t$  are:

$$d[0]/dt = -k_{01}[0] - k_{02}[0] + k_{10}[1] \quad (1)$$

$$d[1]/dt = k_{01}[0] - k_{10}[1] - k_{12}[1] \quad (2)$$

$$d[2]/dt = k_{02}[0] + k_{12}[1] \quad (3)$$

Where each rate constant is defined by the following parameters ( $\alpha_0$ ,  $\beta_0$ ,  $\gamma_0$  and  $\delta_0$  are coefficients used to calculate each rate constant as a function of force and [Trx]):

$$k_{01} = \alpha_0[\text{Trx}] \quad (4)$$

$$k_{12} = \beta_0 \exp(F\Delta x_{12}/k_B T) \quad (5)$$

$$k_{02} = \gamma_0[\text{Trx}] \exp(F\Delta x_{02}/k_B T) \quad (6)$$

$$k_{10} = \delta_0 \quad (7)$$

As shown in the kinetic model scheme in Supplementary Fig. 4a (also shown in Fig. 2d), rate constants  $k_{01}$  and  $k_{02}$  are linearly dependent on the concentration of Trx and have units of  $\mu\text{M}^{-1} \text{s}^{-1}$ .  $k_{10}$  is a constant with units of  $\text{s}^{-1}$ .  $k_{12}$  is in units of  $\text{s}^{-1}$  and is modelled to be exponentially dependent on the applied force, following the Bell equation<sup>15</sup>.  $k_{02}$  also demonstrates an exponential dependence on the applied force. There are no reverse rate constants for the 0→2 and 1→2 transitions (that is,  $k_{20} = 0$  and  $k_{21} = 0$ ). Immediately after disulphide bond reduction in an I27<sub>SS</sub> module by Trx, the two thiol groups in I27<sub>SS</sub> are pulled more than 10 nm apart by the applied force. This prevents any reoxidation of the disulphide bond in I27<sub>SS</sub>, so the formation of state 2 is irreversible in our experiment. We assume that the concentration of Trx remains constant throughout the experiment because the rare oxidation of single enzymes will not significantly affect the overall solution concentration of active Trx. The [Trx] term is input as a constant from the experimental conditions. The I27<sub>SS</sub> concentration is not a factor in the rate equations because only single molecules are monitored at any given time and all results are unaffected by the bulk concentration of I27<sub>SS</sub>.

To describe the obtained experimental data, we solved this kinetic model using matrix analysis. By determining the eigenvalues and eigenvectors of the kinetic matrix (see equation (8)) it is possible to calculate the probability of a single I27<sub>SS</sub> module being in a given state as a function of time.

$$A = \begin{bmatrix} -(k_{01} + k_{02}) & k_{10} & 0 \\ k_{01} & -(k_{10} + k_{12}) & 0 \\ k_{02} & k_{12} & 0 \end{bmatrix} \quad (8)$$

If we input values for the parameters  $\alpha_0$ ,  $\beta_0$ ,  $\gamma_0$ ,  $\delta_0$ ,  $\Delta x_{12}$  and  $\Delta x_{02}$  as well as the experimental [Trx], we can solve the matrix for a discrete set of forces in the range of 0–600 pN. The output of the analysis shows the probability of a single disulphide bond existing in state 0 (Supplementary Fig. 4b), state 1 (Supplementary Fig. 4c) or state 2 (Supplementary Fig. 4d) as a function of time. We note that the model is solved with the initial condition of  $P(0) = 1$  at



time = 0. State 2, where disulphide reduction by Trx has occurred, is the only state that we can directly monitor using our experimental technique. Thus, the calculated probability of being in state 2 as a function of time directly corresponds to the observed single-molecule recordings shown in Fig. 2a. By fitting these calculated probabilities with a single exponential (Supplementary Fig. 4d), we can obtain the observed rate constant of reduction ( $r = 1/\tau$  from the exponential fit to the model plot) in the same manner that we determined  $r$  for the experimental data.

To find the optimal kinetic parameters to describe the experimental data, we first solved the kinetic model for several, widely ranging values for each parameter (typically over three orders of magnitude). We then compared the model  $r$  values to those obtained experimentally (see Fig. 2c, d for wild-type Trx, and Fig. 3b for Trx(P34H)) and calculated the goodness of the fit  $\chi^2$ , where  $\chi^2 = \sum_{i=1}^N \left( \frac{y_i - f(x_i)}{\sigma} \right)^2$ ; here,  $N$  is the number of data points,  $y_i$  is the experimentally observed rate,  $f(x_i)$  is the calculated rate from the kinetic model, and  $\sigma$  is the magnitude of the error of the observed rate<sup>35</sup>. The combination of parameter values with the lowest  $\chi^2$  then served as the starting point for the downhill simplex method<sup>36</sup> to optimize further the global fit of the model to the data.

Errors for each parameter were again obtained with the bootstrap method in combination with the downhill simplex method. At each given value of force and [Trx], a value for the rate constant from the distribution obtained with the bootstrap method (see the 'Data Analysis' section above) was picked at random. By using different values for the rate at each force, extracted from the bootstrap analysis, the experimental error in each rate constant is accounted for when performing fits to the model. The downhill simplex method was then applied to these rate constants, giving the best fitting values for each parameter for that particular combination of rates. The downhill simplex simultaneously varied all six fitting parameters to globally fit the 32 data points for wild-type Trx (14 at 8  $\mu$ M Trx; 18 at other concentrations). This procedure was repeated 200 times, resulting in distributions, and thereby standard errors, for each model parameter. The values that provided the best fit to the data as well as their standard errors for wild-type Trx, Trx(P34H) and TRX are shown in Supplementary Table 1.

To determine the goodness of fit of the various kinetic models shown in Supplementary Fig. 5, we first used the above methods to globally fit each model to the force-dependent and concentration-dependent data for wild-type Trx (Supplementary Fig. 6). We measured an overall  $\chi^2$  value for the best fit to each model (best-fitting parameters shown in Supplementary Table 2). We then obtained a reduced chi-squared value,  $\chi^2_v = \chi^2/v$ , where  $v$  is the number of degrees of freedom in the fit ( $v = N - c$ , where  $N$  is the number of data points and  $c$  is the number of free fitting parameters). To determine the statistical goodness of fit, we calculated  $P(\chi^2_v)$ , the likelihood of obtaining the observed  $\chi^2_v$  if the experimental data are truly represented by the proposed kinetic model<sup>35</sup>. This method has been used previously to determine the goodness of fits of various kinetic models to single-molecule data<sup>12</sup>.  $P(\chi^2_v)$  was calculated using the web-based program available at <http://www.fourmilab.ch/rpkp/experiments/analysis/chiCalc.html>. Parameters relating to the analysis of various kinetic models are shown in Supplementary Table 3.

**Force-probe molecular dynamics and structural modelling.** Simulations were carried out with the Gromacs 3.3.1 simulation suite (<http://www.gromacs.org>)<sup>37</sup>. The simulations were started from the NMR structure of human TRX in an intermediate complex with a disulphide bond to a substrate, the NF- $\kappa$ B peptide (PDB accession number 1MDI, ref. 18). Protonation states of the standard amino acids were adopted from the solution structure.

The OPLS (optimized potentials in liquid simulations) force field<sup>38</sup> was applied. The protein was solvated in a  $7.3 \times 7.3 \times 7.4$  nm<sup>3</sup> box of TIP4P water molecules. Twenty-two sodium and 18 chloride ions were added to the simulation system to compensate for the overall positive charge of the protein and to mimic physiological conditions. This yielded a total system size of 49,220 atoms. Simulations were carried out with periodic boundary conditions. Application of the Lincs<sup>39</sup> and Settle<sup>40</sup> methods allowed for an integration time step of 2 fs. Electrostatic and Lennard-Jones interactions were calculated within a cut-off of 1 nm, and the neighbour list was updated every ten steps. For the long-range electrostatic interactions, the Particle-Mesh-Ewald (PME) method<sup>41</sup> with a grid spacing of 0.12 nm was used. An  $N, p, T$  ensemble, where  $N$  is the number of atoms,  $p$  is the pressure and  $T$  is the temperature, was simulated, with separate coupling of the protein, solvent and ions to a 300 K heat bath ( $\tau = 0.1$ , ref. 42). The system was isotropically coupled to a 1 bar pressure bath ( $\tau = 1.0$ , ref. 42). Initially, the system was energy-minimized (steepest descent, 1,000 steps), before equilibrating the solvent for 700 ps with positional restraints on protein heavy atoms. Then, the whole system was equilibrated (300 K). Input coordinate files and system parameters are also included in the Supplementary Information.

(Owing to the large size of the output file it is not practical to include it in the Supplementary Information. Therefore, we will send the output file of our simulations to any researcher, on request.)

To model an approximate transition state geometry for the  $S_N2$  reaction in the active site of Trx, in a subsequent simulation the Trx–NF- $\kappa$ B disulphide bond was elongated from 2.05 Å to 2.60 Å (the length of the extended bond found for the transition state in an  $S_N2$  reaction<sup>22</sup>) within 160 ps using the free-energy perturbation code in Gromacs and starting from the equilibrated system. Next, the third sulphur atom taking part in the  $S_N2$  reaction was placed along the resulting vector of the extended disulphide bond between Trx and the NF- $\kappa$ B peptide in a distance of 2.40 Å, as found for the  $S_N2$  transition state<sup>22</sup>. The cysteine residue to which the third sulphur atom is bound was placed into the location defined by the sulphur atom, and was oriented such that it did not clash with Trx or peptide residues. Using 20 different starting structures of equilibrated Trx for the modelling of the reduction transition state resulted in somewhat different active-site geometries. The angle between the peptide-binding groove and the axis of the sulphur atoms varies and exists in the range between 50° and 130°. The average conformation of the disulphide bond was observed to fall into two populations (Supplementary Fig. 10). The resulting structures were plotted with Pymol<sup>43</sup>.

In another set of simulations, titin I27 with residues 32 and 75 mutated to cysteines was unfolded to monitor the disulphide bond orientation in the unfolded state with respect to the pulling direction. The OPLS force field was applied for I27. The wild-type protein (PDB accession number 1TIT, ref. 44) was solvated in TIP4P water in a  $6.8 \times 5.7 \times 5.0$  nm<sup>3</sup> box. Sixteen sodium and ten chloride atoms were added to neutralize the protein charges and to give physiological ion strength. The resulting system size was 23,524 atoms. I27 was minimized, the solvent initially equilibrated with restraints on the protein heavy atoms (500 ps), and then the entire system subsequently equilibrated for a further 8 ns. The simulation software and parameters as described above were applied. Residues 32 and 75 of the equilibrated structure were mutated to cysteine residues using the program WHATIF (ref. 45). The mutant I27<sub>SS</sub> was re-solvated in a larger box ( $19.2 \times 5.5 \times 5.0$  nm<sup>3</sup>), allowing sufficient space to completely unfold the protein, yielding a system size of 112,156 atoms. The system was minimized, resulting in a shortening of the S–S bond to the value typical for an S–S bond (2.05 Å). The solvent was equilibrated with restraints on the protein heavy atoms (2 ns), followed by the equilibration of side chains with restraints on the protein backbone atoms (2 ns), and finally by the equilibration of the whole system (11 ns). No distortion of the structure adjacent to the point mutations was observed. Force-probe molecular dynamics simulations<sup>46</sup> of the equilibrated I27<sub>SS</sub> mutant were performed. The  $C_\alpha$ -atoms of the terminal residues were subjected to harmonic pulling potentials with a spring constant of 500 kJ mol<sup>−1</sup> nm<sup>−2</sup>, and were moved away from each other with a constant velocity of 0.4 nm ns<sup>−1</sup>. As expected, the unfolded structure (Supplementary Fig. 8), obtained after ~14 ns of the force-probe molecular dynamics simulation time, showed alignment of the disulphide bond within ~20° of the pulling direction, with a projection of the S–S bond length on the pulling axis of ~1.9 Å.

For comparison of the active-site geometry, additional standard equilibrium molecular dynamics simulations have been performed for the reduced state of Trx in the absence of a peptide, and for the other available Trx intermediate, the Trx–Ref-1 complex. For the simulation of the reduced state, the NF- $\kappa$ B peptide in the 1MDI structure was deleted. The apo structure with an unprotonated Cys 32 was solvated in water. After addition of ions to yield physiological ion strength, the system comprised 33,606 atoms. The Trx–Ref-1 complex (PDB accession number 1CQH, ref. 19) was solvated in water with physiological ion strength, resulting in a system size of 38,760 atoms. Both systems were minimized and equilibrated as described above and results are shown in Supplementary Fig. 10.

26. Perez-Jimenez, R., Godoy-Ruiz, R., Ibarra-Molero, B. & Sanchez-Ruiz, J. M. The effect of charge-introduction mutations on *E. coli* thioredoxin stability. *Biophys. Chem.* **115**, 105–107 (2005).
27. Holmgren, A. & Reichard, P. Thioredoxin 2: cleavage with cyanogen bromide. *Eur. J. Biochem.* **2**, 187–196 (1967).
28. Holmgren, A. & Slaby, I. Thioredoxin-C': mechanism of noncovalent complementation and reactions of the refolded complex and the active site containing fragment with thioredoxin reductase. *Biochemistry* **18**, 5591–5599 (1979).
29. Ren, X., Bjornstedt, M., Shen, B., Ericson, M. L. & Holmgren, A. Mutagenesis of structural half-cystine residues in human thioredoxin and effects on the regulation of activity by selenodiglutathione. *Biochemistry* **32**, 9701–9708 (1993).
30. Fernandez, J. M. & Li, H. Force-clamp spectroscopy monitors the folding trajectory of a single protein. *Science* **303**, 1674–1678 (2004).
31. Oberhauser, A. F., Hansma, P. K., Carrion-Vazquez, M. & Fernandez, J. M. Stepwise unfolding of titin under force-clamp atomic force microscopy. *Proc. Natl Acad. Sci. USA* **98**, 468–472 (2001).

32. Sigworth, F. J. & Neher, E. Single Na<sup>+</sup> channel currents observed in cultured rat muscle cells. *Nature* **287**, 447–449 (1980).
33. Aldrich, R. W., Corey, D. P. & Stevens, C. F. A reinterpretation of mammalian sodium channel gating based on single channel recording. *Nature* **306**, 436–441 (1983).
34. Efron, B. *The Jackknife, the Bootstrap, and Other Resampling Plans* (SIAM, Philadelphia, 1982).
35. Taylor, J. R. *An Introduction to Error Analysis* 2nd edn 261–293 (Univ. Science Books, Sausalito, 1997).
36. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes in C* (Cambridge Univ. Press, Cambridge, UK, 1992).
37. Lindahl, E., Hess, B. & van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* **7**, 306–317 (2001).
38. Jorgensen, W. L. & Swenson, C. J. Optimized intermolecular potential functions for amides and peptides. Structure and properties of liquid amides. *J. Am. Chem. Soc.* **107**, 569–578 (1985).
39. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
40. Miyamoto, S. & Kollman, P. A. Settle—an analytical version of the shake and rattle algorithm for rigid water models. *J. Comput. Chem.* **13**, 952–962 (1992).
41. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald—an  $N \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
42. Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
43. Delano, W. L. *Pymol Manual* (<http://www.delanoscientific.com>) (2001).
44. Improt, S., Politou, A. S. & Pastore, A. Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity. *Structure* **4**, 323–337 (1996).
45. Vriend, G. WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.* **8**, 52–58 (1990).
46. Grubmüller, H., Heymann, B. & Tavan, P. Ligand binding: molecular mechanics calculation of the streptavidin biotin rupture force. *Science* **271**, 997–999 (1996).

## ERRATUM

doi:10.1038/nature06362

**Insight foreword: Food and the future**

H. Gee

*Nature* **418**, 667 (2002); doi:10.1038/418667a.

This article was initially assigned the digital object identifier (DOI) number 10.1038/nature01012. Subsequently, the same DOI number was inadvertently assigned to another article<sup>1</sup>. A new DOI number—10.1038/418667a—has now been assigned to this article and should be used when citing it.

1. Roy, P. J., Stuart, J. M., Lund, J. & Kim, S. K. Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*. *Nature* **418**, 975–979; doi: 10.1038/nature01012 (2002).



# naturejobs

**JOBS OF  
THE WEEK**

Cases of misconduct in science, ranging from unearned authorship to falsifying data, seem to crop up with surprising regularity. Every effort needs to be made to ensure that fledgling researchers understand what is — and isn't — acceptable. For postdocs, such training is especially important because they are at a point in their career at which they are still being mentored but are also responsible for mentoring others.

In particular, as the US National Postdoctoral Association (NPA) notes, postdocs often have no formal standing at their institution, which means they may have to rely on their overworked — or perhaps sometimes unreliable or even untrustworthy — mentor to address issues such as who should be first author on a paper, how or when collaborations should take place, and what happens to the postdoc's data when he or she leaves the lab. These are potentially awkward topics to broach, and many postdocs avoid them, allowing the daily research tasks to take precedence. But they are also issues in which research integrity could become compromised.

To address the problem, the NPA launched a programme to increase awareness and promote training in the responsible conduct of research. Last month it announced the first round of seed grants in the programme. Twelve universities and postdoc associations each received \$1,000 to help them with initiatives centred on research integrity. The Massachusetts General Hospital Postdoctoral Association in Boston, for example, will use its money to support a series of lunch discussions between mentors and postdocs. At the University of Pittsburgh in Pennsylvania, the grant is helping to catalyse conferences on various topics related to 'responsible conduct'.

There is always the danger that such discussions will do little more than pay lip service to the issues and offer little additional practical advice. But even in this worst case, they will achieve something highly important — they will get people talking about these problematic areas. And that could be invaluable, because if this ice isn't broken, issues such as determining authorship or setting up a collaboration could become as complex and challenging as the science itself.

**Gene Russo, acting editor of *Naturejobs***

## CONTACTS

**Acting Editor:** Gene Russo

### European Head Office, London

The Macmillan Building,  
4 Crinan Street,  
London N1 9XW, UK  
Tel: +44 (0) 20 7843 4961  
Fax: +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

### European Sales Manager:

Andy Douglas (4975)  
e-mail: [a.douglas@nature.com](mailto:a.douglas@nature.com)  
**Business Development  
Manager:**  
Amelie Pequignot (4974)  
e-mail: [a.pequignot@nature.com](mailto:a.pequignot@nature.com)

### Natureevents:

Claudia Paulsen Young  
(+44 (0) 20 7014 4015)  
e-mail: [c.paulsenyoung@nature.com](mailto:c.paulsenyoung@nature.com)

### France/Switzerland/Belgium:

Muriel Lestringuez (4994)

### Southwest UK/RoW:

Nils Moeller (4953)

### Scandinavia/Spain/Portugal/Italy:

Evelina Rubio-Hakansson (4973)

### Northeast UK/Ireland:

Matthew Ward (+44 (0) 20 7014 4059)

### North Germany/The Netherlands:

Reya Silao (4970)

### South Germany/Austria:

Hildi Rowland (+44 (0) 20 7014 4084)

### Advertising Production Manager:

Stephen Russell  
To send materials use London  
address above.

Tel: +44 (0) 20 7843 4816

Fax: +44 (0) 20 7843 4996

e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

### Naturejobs web development:

Tom Hancock

### Naturejobs online production:

Jasmine Myer

### US Head Office, New York

75 Varick Street, 9th Floor,  
New York, NY 10013-1917  
Tel: +1 800 989 7718  
Fax: +1 800 989 7103  
e-mail: [naturejobs@natureny.com](mailto:naturejobs@natureny.com)

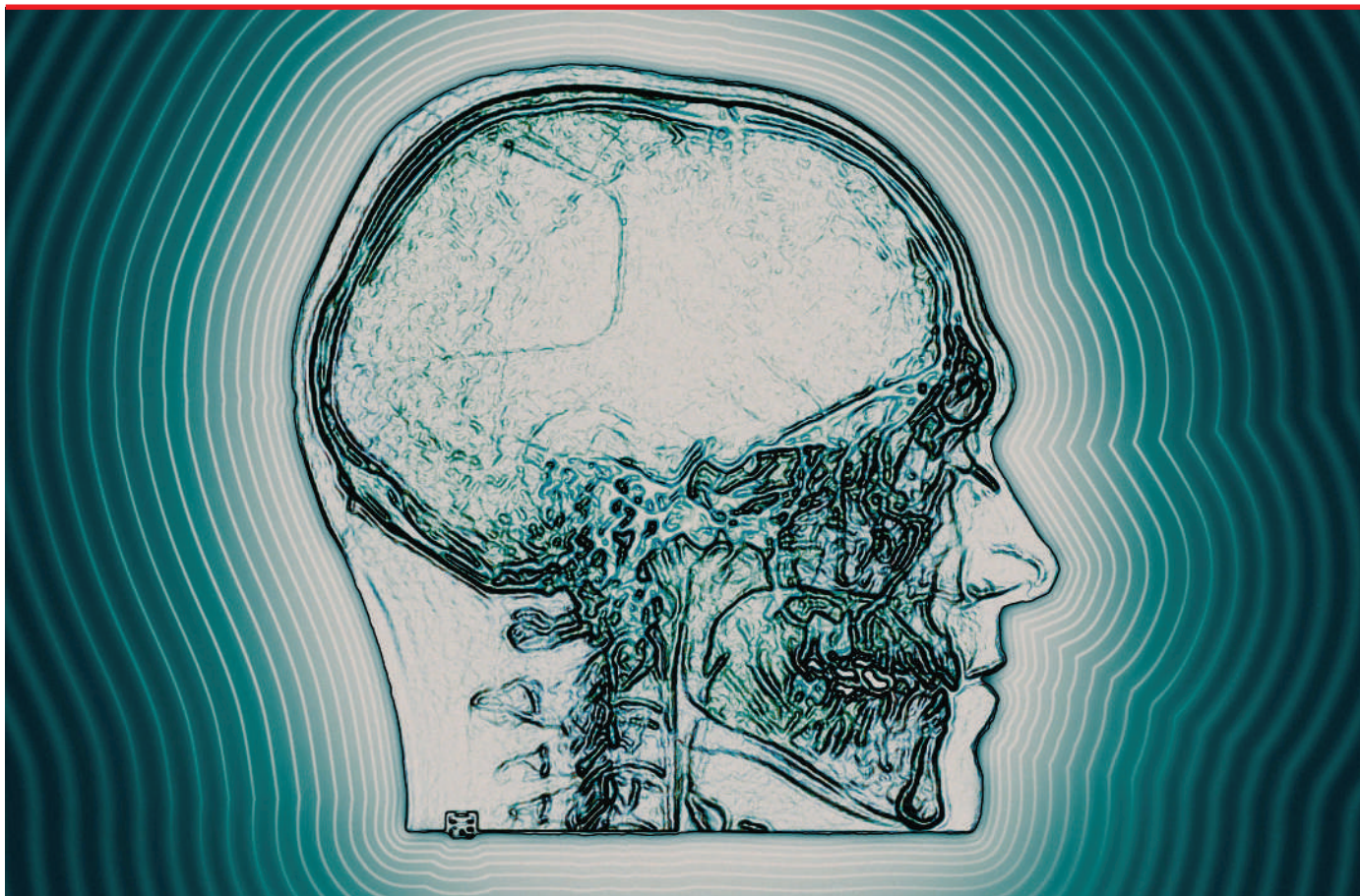
### US Sales Manager: Peter Bless

### Japan Head Office, Tokyo

Chiyoda Building,  
2-37 Ichigayatamachi,  
Shinjuku-ku, Tokyo 162-0843  
Tel: +81 3 3267 8751  
Fax: +81 3 3267 8746

### Asia-Pacific Sales Manager:

Ayako Watanabe  
Tel: +81-3-3267-8765  
e-mail: [a.watanabe@natureasia.com](mailto:a.watanabe@natureasia.com)



# BRAIN STORM

A bold scheme to map the entire human brain has become the mission of many scientists from a host of different fields. **Paul Smaglik** tracks the interdisciplinary career implications.

**A** major goal for neuroscientists is to build a complete wiring diagram of the human brain. Such a feat has already been achieved for the nematode worm *Caenorhabditis elegans*. But going from worm to human requires a serious leap: *C. elegans* has 302 neurons connected together through 7,000 synapses; the human brain has an estimated 10 billion neurons, each of which has an average of 10,000 synaptic connections.

Undaunted, many neuroscientists are pursuing at least part of the problem, and funding is following. But with no centralized programme for this effort, there are multiple approaches, ranging from the use of electrophysiology to tease out synaptic connections in specific locations, to building mathematical models of neural networks. As yet there is no consensus on which approach is best, but there is general agreement that building and using different technologies — from the single-cell focus of electrophysiology to the whole brain views captured by functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) — will bring a brain map closer to reality.

Today, neuroimaging and computational neuroscience are akin to genomics in the 1980s, with many groups working in their own niches. Physicists are building better imaging tools; computational experts and statisticians are pondering theoretical frameworks; and neuroscientists are trying new data-

gathering methods. The sheer scope means opportunities for scientists from many disciplinary backgrounds. But there is also some professional risk: the field as a whole remains uncertain about whether any single imaging or computational approach can map the trillions of connections that make up the brain's neural network; and the people who build and use the technological tools involved may not be rewarded as directly as those who interpret the data.

Perhaps the biggest challenge will be combining myriad images from sources such as electron microscopy, PET and fMRI, says David Van Essen, president of the Society for Neuroscience. Van Essen organized a series of lectures at the society's meeting this week in San Diego to explore this issue and show some examples of how this is being done. He believes that reluctance to adopt an approach that combines multiple kinds of neural imaging with computational techniques goes beyond scientific reasons. "There are challenges to getting integrative and interdisciplinary projects like this funded in a very competitive environment," he says.

Large-scale, team-based approaches should help align images from different sources on top of each other — from single synapse to large-scale data using PET and fMRI. This should help elucidate how the behaviour of individual cells affects broader brain activity.

Scientists — from engineers who build better



**Nikos Logothetis is planning data analysis using lots of mathematics.**



ADAPTED WITH PERMISSION FROM VAN ESSEN AND DIERKER (NEURON 2007)

imaging machines to statisticians who will help make sense of the data — will be invaluable. But first, scientists need a better theoretical framework, says Lyle Graham, who leads a group in visual neurophysics and physiology, at France's basic-research agency, the CNRS, in Paris.

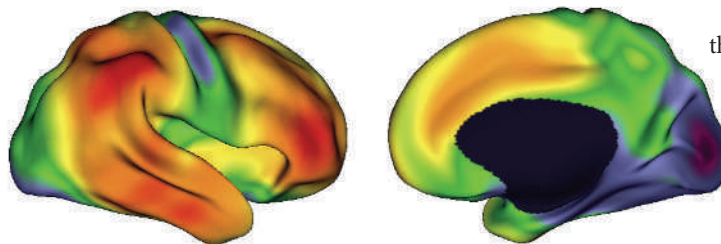
Graham proposes combining multiple approaches, starting with electrophysiology and computational modelling. Physicists, mathematicians and engineers need to appreciate the nuts and bolts of theory-driven experiments, he notes, and biologists need to understand the role of the computational approach.

Graham says it is not easy recruiting the sort of good experimentalists who are essential for testing hypotheses of brain circuitry and function. "Molecular biology is much more attractive to people doing their PhDs or postdocs," he says. "In electrophysiology, you're not sure it's going to work. It's hard. You have to be a little obsessive." He recommends that people who are considering the field first ground themselves in both the biophysical bench approaches and the theoretical models of how the brain works, so that theory and experiment can inform each other.

### All in the mind

Liam Paninski, a statistician at Columbia University in New York who works on neural codes, is one such person. As an undergraduate, he participated in single-synapse research, but soon discovered he needed a better grasp of statistics to model and validate the data. Although his PhD is in neuroscience, he emphasized statistical methodology as part of his training. That background is helping him make the leap from using electrodes — which provide high-resolution data, but are also highly invasive as they have to be inserted into the living brain — to calcium-sensitive dyes that show electrical activity in the synapses. He is using statistics to 'clean up' the calcium data and to extrapolate them to multiple cells. "You can dump calcium into the brain and look at a whole region," he says.

Paninski believes that multidisciplinary support for learning about the brain's connections is growing. The US National Institutes of Health and National Science Foundation sponsor a grant for collaborative research in computational neuroscience, to bring mathematicians, computational experts and physiologists together. "When I was starting out ten years ago I thought I had to run my own lab and collect my own data," he says. Now he works with a team of neuroscientists and physicists who help him collect the data that he



Combined data have been used to create these two views of cortical expansion between monkey and humans. Red patches indicate regions that are likely to have expanded most rapidly in humans.

then analyses statistically.

Winfried Denk of the Max Planck Institute for Medical Research in Heidelberg, Germany, says one approach to building a map is to start small, imaging all the neurons in a very limited section of brain

tissue using electron microscopy. "We're trying to develop a method to basically reconstruct detailed circuit diagrams of neural tissues," Denk says. But he suggests that with existing approaches it would take "a billion people a year" to capture all of the synapses in the cortical column alone. "That's not quite practical," he says. Instead, Denk is working with Sebastian Seung's lab at the Massachusetts Institute of Technology in Cambridge. Seung is helping him to generate algorithms that show the result of many synaptic connections, rather than imaging every single one.

### Sexy genes

Heidi Johansen-Berg is using diffusion fMRI at the John Radcliffe Hospital in Oxford, UK, to study anatomical connections in the brain. Although it supplies less detail than electron microscopy or electrophysiology, it can still help build a lower-resolution wiring diagram, she says. Such data can then be compared to the human genome and the Allen Brain Atlas, which has mapped out the expression patterns of more than 20,000 genes in the mouse brain.

She anticipates more convergences in imaging and data analysis to come. "It's a very exciting time to enter the field," Johansen-Berg says. "A challenge will be to see how taking a computational approach will help in identifying and describing disease state."

Interdisciplinary groups are trying to combine imaging approaches and analyse them with statistics, and computational and mathematical modelling. Nikos Logothetis, a professor in physiology and cognitive processes at the Max Planck Institute for Biological Cybernetics in Tübingen, Germany, has combined data from fMRI and electrophysiology and wants to do both approaches simultaneously, then use "lots and lots of mathematics" to build circuit diagrams.

Karel Svoboda, a group leader at the Howard Hughes Medical Institute's Janelia Farm campus, says that all 15 neuroscience groups there take a similar approach, but concentrate on specific regions of the cerebral cortex. Janelia plans to add 10–15 more leaders, says Svoboda: investigators examining neural circuits who use imaging tools; physicists, engineers and computational scientists who are building tools; and "eclectic others, who are smart and clever and fun to have around".

To "reverse engineer the brain", more institutes and universities will need to build small multidisciplinary teams like Janelia's, Svoboda says. "This field needs an influx of engineers and applied physicists," he says. Mapping brain structure and function is a broader, deeper problem than sequencing and analysing the genome, he says. For now, that goal will provide plenty of work for multidisciplinary-minded scientists — and could create even greater opportunities if and when the drive to draw up the wiring diagram of the human brain gets the kind of leadership and funding afforded to the Human Genome Project.

**Paul Smaglik is a freelance writer based in Milwaukee, Wisconsin.**



Multitasking: (clockwise from top) Liam Paninski, David Van Essen and Lyle Graham.

R. SUSSMAN

CNRS, UNIVERSITÉ PARIS DESCARTES



# MOVERS

**Vasilis Ntziachristos, director, Institute for Biological and Medical Imaging, GSF, Neuherberg, Germany**



**2002-07:** Assistant professor, Harvard University Medical School and Massachusetts General Hospital, Boston, Massachusetts

**2001-02:** Instructor, University Medical School and Massachusetts General Hospital, Boston, Massachusetts

As an engineering undergraduate at the Aristotle University of Thessaloniki in his native Greece, Vasilis Ntziachristos was attracted to biomedical imaging as a way to use his skills to benefit others.

For his undergraduate thesis, he developed hardware and software platforms to improve the performance of magnetic resonance imaging (MRI). To get hands-on experience with imaging applications, Ntziachristos took a fellowship at the Panum Institute at the University of Copenhagen in Denmark. "It soon became apparent to me that I wanted to live at this interface between engineering, medicine and biology," he says.

One year later, he was a graduate student at the University of Pennsylvania in Philadelphia working with leading imaging scientist Britton Chance. There, Ntziachristos combined MRI with the emerging power of optical imaging to determine the molecular structure and function of tissues. The result was a new imaging approach to determine the molecular characteristics of breast cancer.

As a research fellow at Harvard Medical School in Boston, Ntziachristos continued to refine the optical technique to identify the molecular signatures of various tissues. Taking advantage of the concurrent explosion of genomic information, he developed techniques to decipher the molecular signatures of genes and proteins *in vivo*.

After seven years at Harvard, Ntziachristos is now moving on to a post in Germany. In November, he will take over as director of the Institute for Biological and Medical Imaging at the GSF-National Research Center for Environment and Health and will also become chair for the biological-imaging department at the Technical University of Munich. "Vasilis will help us visualize molecular events to better understand biology and pathology," says Gunther Wess, president of the GSF.

Ntziachristos has two major goals for the institute. He wants to create significantly better biological tools, and to move those tools into the clinic so that doctors can make more informed treatment decisions. His first challenge is building the multidisciplinary environment necessary for this work — he needs everything from chemistry expertise to skills in working with animal models — under one roof. Wess says it was Ntziachristos's collaborative zeal, as well as his technical expertise, that made him the perfect fit for such ambitions at the GSF.

**Virginia Gewin**

## NETWORKS & SUPPORT

### Brain work

Jeff Hawkins got the neuroscience bug in 1979, but it took him many years and two inventions — the Palm Pilot and the Treo smartphone — to put his passion into action. His career arc illustrates how technological acumen and a theoretical mindset can foster one another: both his devices use pattern matching and computational decision-making algorithms that mimic theories of how a brain works.

As an undergraduate at Cornell University, Hawkins was impressed with a *Scientific American* article by Nobel laureate Francis Crick. "He was saying we have all these data about neurons and chemical processes in the brain and people act like they know what's going on, but don't believe it," Hawkins says. "What's lacking is a theoretical framework." Models were needed to show how the cells and processes interacted.

Hawkins suggested starting a neuroscience division at his first employer, Intel, but was rebuffed. He applied some of his theories about pattern recognition at Grid Systems in Sunnyvale, California, where he worked on handwriting recognition hardware and software. To learn more about how the brain processes pattern-matching, Hawkins went to the University of California, Berkeley, where he researched the workings of

the brain's neocortex, which is associated with vision and decision-making. But he was told neocortical theory wasn't a viable thesis option.

That gave him insight into the obstacles of trying to develop an unfashionable field. "I learned that understanding the brain is not just a scientific problem, it's an institutional problem," Hawkins says.

His next few years were spent back at Grid, which later became Palm, developing Palm Pilot and Treo — which he calls "sort of my day job, like an actor waiting on tables".

His success led him to start the Redwood Neuroscience Institute in Menlo Park, California, and co-write *On Intelligence* (Times Books, 2004), which expands on themes from his rejected thesis proposal. He later formed Numenta, which designs three-dimensional decision-making trees that can be applied to things from motion-capture video games to traffic sensors mounted on vehicles.

Hawkins thinks applications may drive more institutional change and private investment into neuroscience. "The commercial world is a very powerful way to get people to work on a problem," Hawkins says. "If you can show a profit, you can get a lot of people to work on this."

**Paul Smaglik**

#### POSTDOC JOURNAL

### Corporate America

I recently contacted a former colleague who is now working in agro-industry. I am considering a career in industry, and I wanted to pick her brain for advice. She has worked in her position as group research leader for only a year, but she had some helpful insights. Although she has the job she wanted, the number of management duties has reduced her self-image as a serious scientist. Also, as she and her husband aren't able to live in the same state, it's nearly impossible for them to start a family. They've had trouble deciding who should give up their job to allow them to live together, even though she admits that this isn't her permanent career path.

That said, it's not all bad. Her company believes that leadership and scientific collaboration are important and encourages staff to attend training sessions, conferences and collaborative projects. Also, the pay is good, and the facilities help projects go faster and more smoothly than they might at a university.

Although not all research lends itself to a corporate output, I've come to realize that for many this is a good option. In particular, I like the way companies tend to be more family-friendly than universities — enabling employees to keep more regular hours, for example. For me, the hardest adjustment is the notion that science is profit, and that this has great influence on one's research.

**Maira Sheehan is a postdoc in the Department of Plant Breeding and Genetics at Cornell University.**

# A sudden absence of bees

You've been stung.

**Nick Mamatas**

Enrolment was down. Even the girls with hair down to their hips and lazy smiles know better than to major in literature any more, so the chair had me develop a course to get butts on seats. I took a multidisciplinary approach — English, pol sci, history — to come up with “(Under)mining Consensus Reality: Conspiracy Theories”. We read *The Da Vinci Code* and *Holocaust on Trial*, watched *JFK*, and then the students did a creative final — writing their own conspiracy theories.

I got lots of 9/11 from junior leftists; man-made AIDS from the militant queers and Afrocentric kids; and comical suppositions about UFOs and *American Idol* from the rest. Except the Front Row: Kap, Julia and Aram. These three bio majors hadn't been in the humanities building since frosh comp, but now they needed a writing-intensive class to graduate and sat together, all quiet intensity and perfectly formatted term papers.

Kap was born in Seoul. In the battle between his adoptive born-again Christian parents and his craving for a seat in Stony Brook's med school, he lost. He wrote “Toward A Postmodern Materialist Buddhism”. There is an Intelligent Designer, he argued, despite the fact that humans are remarkably poorly designed. But the Designer lacks experience in limitation and pain. The ID could only experience limitation vicariously, so he makes humanity full of bursting appendices, posterior superior parietal lobes hardwired for religious experience, and with expiration dates. ID theory is sound once one realizes that the ID is a CT: a Cunning Torturer. I gave Kap a B<sup>+</sup>: conspiracy theories involve human agency, not a Prime Mover.

Julia was into ecology. Like many outdoorsy girls, she'd developed a great wit to offset her boat hips and mannish chin. “Why Blondes Have More Fun”, she wrote. She also got a B<sup>+</sup>, mainly because she obviously wrote the paper for a bio class, and then tacked on a page of ‘natural’ conspiracy. A wasp, *Hymenopimecis* sp. conspires against the *Plesiometa argyra* spider. The wasp attacks and lays its eggs within the spider's abdomen. The larva consumes the spider's haemolymph (“spider blood”, she explained) and then excretes a chemical that changes the spider's behav-

iour. Instead of the web *Plesiometa argyra* usually builds, the chemical compels the spider to create a box-web design that can support the weight of the pupating wasp. After the web is done, the larva eats the spider, builds a cocoon in the strong web and then pupates.

On the last page of the paper Julia explained that natural blondes secrete a similar chemical when they reach men-ses. It makes men take blonde girls out to



dinner and then, later in life, provide them with houses and cars. Then the men die of early heart attacks due to all the fancy steaks they consumed and the blondes keep the swag. Cute, but I know when a junior is phoning it in.

Aram's theory was “Beepocalypse Now”. Producers of genetically modified seeds, he claimed, could create a plant whose pollen would act as a soporific for bees. Rather than returning to their hives, the bees, “dazed and confused for so long” (Aram, all acne and black concert T-shirts, was really into Zep), would just wander off and die. The ecosystem would collapse, allowing agribusiness to monopolize production with GMO crops, which don't need pollination ... for a price. He got a B<sup>+</sup> as well. The assignment was to write a current event, not a speculative one.

The Front Row weren't happy with their grades, nor with the reasoning I presented during office hours. Then, 18 months later, the bees began to vanish.

This was followed by a rash of church

burnings sweeping the United States, and then much of the rest of the world. This wasn't a sectarian strike either; houses of worship of all faiths were targeted. Focus on the Family shut its doors, and in Kansas several former school board members hanged themselves. Pakistan and India began reunification talks and Saudi Arabia witnessed its first ever Pride march, although it ended in violence. Everything ends in violence now.

The kids started demanding more coursework on Nietzsche, and their interpretations of both canonical and postmodern literature got decidedly darker. I drank my way through long nights of grading papers — college writing is almost always bad, but it had never been so twisted. Religiosity hadn't disappeared; instead it warped into a sort of irrational existential angst. Crime and suicide rates skyrocketed. After the latest mass shooting on a college campus, the AAUP demanded an increase in budgeting for distance learning purely for faculty safety.

I realized what was going on while huddled under my desk. The Front Row had done it. The alumni newsletter confirmed my theory. Aram was working on GM crops for the United Nations, and Julia was ‘in the field’ trying to find the wayward bees. Her ‘theory’ was that *Hymenopimecis* sp. had moved into new ecological niches and crowded them out. Kap was head of a new online faith — the First Virtual Church of Christ the Deleted — that was gaining converts as fast as the world's religions were losing them. They'd proved me wrong.

They couldn't have combined their theories into one bloody actuality just to show me up, could they? That is conspiracy's fallacious attraction. Small cabals don't target history or human events, but theories target the *Weltanschauung* of their audiences. To deal with the cognitive stress of living we write soothing stories that return us to the centre of the Universe. A malevolent universe is as consoling as a beneficent one. And there is a wasp, the first I've seen in years, trapped between panes of glass in my office window. I don't know how it got there, but I know that I must know. I lift the glass and invite it to sting.

**Nick Mamatas is the author of *Under My Roof*, a novel of neighbourhood nuclear proliferation for children, and other books and stories. He lives near, but not in, Boston, Massachusetts.**

JACEY